

AD-A277 847

REF

Agency is The Adjutant General's Office.

MAR 28 1994

REI



DAPE-HRE

Recommendations 3F, Crimes of Violence

TO DAPE-HR

FROM DAPE-HRE

DATE

CMT 1

LTC Barry/bgb/52499

Problem: To study the extent of which the quality of military police is a factor adversely affecting law and order in the Army and the utility of using psychological tests in the MP selection and screening process.

Facts:

1. One of the Department of Army Crimes of Violence Study recommendations approved for further action by the VCSA, on 5 September 1972, was recommendation 3F to study the recommendation reflected in the problem above.
2. Initially an ODCSPER project, the study effort was transferred to OTPMG in a tasking document, 11 July 1973 (Tab A).
3. The tasking document requested OTPMG to form a working group comprised, as a minimum, of a representative from ODCSPER, MILPERCEN, OCRD, TRADOC and USAMPS.
4. The study findings were to be provided DAPE-HRL, NLT 30 Jun 74.

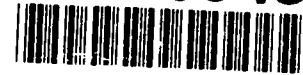
Discussion:

1. After initial coordination with agencies designated to be represented in the working group and discussions with experts in the field of statistical data analysis and psychological testing, the working group concept was discarded in favor of an independent study and periodic coordination. Chief reasons for this decision were shortage of TDY funds, shortage of available personnel to release for a working group, complexity of the data to be researched and analyzed and time available to complete the requirement.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

2. Scope of the study included:
 - a. Review and evaluation of factual evidence of military police exhibiting undesirable behavior, a military police supervisor opinion survey and a community attitude survey on the quality of military police and law enforcement services rendered (Tab B).
 - b. Review of representative work done in the field of psychological testing for police officers over the last 14 years (Tab C).
 - c. Detailed review of studies done by the Industrial Relations Center, University of Chicago (Tab D).
 - d. Summary of the Assessment Center approaches to date in the United States (Tab E).
 - e. Review of the University of Cincinnati Project (Tab F).
 - f. Consideration of the legal aspects of psychological testing (Tab G).



DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

3. The question of what is a quality MP is a relatively complex issue within a military context (Tab B).

a. Despite an apparent understanding of what is meant by quality, a firm useable definition of what is a quality MP could not be found.

b. Without a useable definition of a quality MP for research purposes, the statistical measurement of adverse impact on law and order is impossible.

c. In the absence of a firm statistically workable definition of a quality MP, the study was limited to factual evidence of MP's exhibiting undesirable behavior, MP supervisors' opinions on quality based on perceived job performance requirements, and a survey of community attitudes toward MP's.

4. Three major areas studied on the question of quality MP's revealed:

a. The OTPMG Qualitative Standards Management Program designed to surface evidence of undesirable behavior in MP's thus far has led to the following (Tab B):

(1) Of the 15,227 MP files reviewed to date, some evidence of undesirable behavior has been surfaced.

(2) More than two percent recommended reclassifications represent an insignificant figure in terms of adverse impact on law and order.

(3) More important than the two percent figure are the reasons identified (stress on quantity not quality, lack of lateral entry controls, and failure to promptly reclassify individuals after demonstrated misconduct) which allowed existence of these individuals in the Corps.

b. DAPC-PMP Report No. 17-73E, based on surveys of 20,326 officer and enlisted men conducted by MILPERCEN in the summer of 1972 and 1973, indicates large percentages of the military community surveyed expressed the following (Tab B):

(1) Military police are not effective in a law enforcement/crime prevention role.

(2) The MP qualities on helpfulness and courtesy are below community accepted levels.

c. Residual findings from a report of a team of officers from the Task Analysis/Training Programs Branch, USAMPS, visiting 22 MP units at 17 CONUS installations, revealed that knowledgeable supervisors (i.e. PM's, MP unit commanders and senior enlisted supervisors) felt that (Tab B):

(1) A significant number of USAMPS basic course graduates are too immature to function effectively.

(2) Inadequate selection standards, haphazard screening and low elimination rates for poor achievers in the course result in unqualified MP's serving in CONUS units.

A-1

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

(3) Military police graduates of the basic law enforcement course are not adequately trained in key police tasks.

5. Inherent in problem of utilizing psychological tests in selection of MP's is the existing inability to clearly define tasks and evaluate personnel. Studies of a representation of work done in the field of psychological testing for police officer selection over the past 14 years involved both those dealing with the issue of validity and those focusing on predictions without an attempt to assess validity (Tab C).

a. Validity studies reviewed were all attempts to construct a method to predict police selection success or establish what qualities are required for an individual to perform satisfactory in a police role. Only one study adequately developed job requirements and prediction of police job performance. All but one of the studies reviewed provided, at best, solid evidence for consistent predictive validity between psychological tests of intelligence and aptitude and measures of police academy or scholastic performance. All except one, including the Army classification battery, proved inadequate as a prediction of future police job performance.

b. The only work to date which approximates a useable technique in the selection of MP's are studies of the Industrial Relations Center, University of Chicago (Tab D). Their work covering three projects to date:

- (1) Identifies basic job requirements and required personnel qualifications.
- (2) Selects tests to measure key qualifications.
- (3) Defines job performance criteria.
- (4) Uses paper and pencil tests similar in complexity to present Army tests.
- (5) Attains an unchallenged degree of acceptable validity.

6. Industrial and government organizations have begun to accept the assessment center technique as a valid means to identify and utilize management potential at an early stage. An assessment center is a facility staffed and operated to put candidates through a series of group or individual exercises, interviews and tests designed to stimulate conditions of a supervisory job, and determine if they have skills and abilities necessary to perform the job. For this study, a review was conducted of 25 assessment center techniques from the more than 100 companies and governmental agencies with these programs (Tab E). Observations based on the review are as follow:

a. Assessment centers arrived at objective evaluations of each individual by using a number of simulation problems, administration and interpretation of paper and pencil tests and in-depth interviews.

b. Assessment centers operate with three to four evaluators per six persons evaluated. Evaluate group sizes vary in size between six and 12 persons with the length of sessions varying from one to five days.

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

c. Cost to operate each center varies with the size and complexity of the facility, number of staff employed and administrative requirements. Cost per individual assessed ranges from a low of \$5 to a high of \$500 with the average about \$275.

d. Only one study reviewed addressed selection screening at entry level. The bulk of programs address upward mobility one year or more after entry level; accordingly they do not fit needs of the Army in selection of MP's.

e. The on-going projects developed by Dr. Robert Mills of the University of Cincinnati was the only program tailored solely for law enforcement personnel selection (Tab F). His project is the psychological evaluation phase of the Cincinnati method of police selections. It represents the final step in the screening procedure which includes an AGCT intelligence test, physical examination, individual interview, polygraph and background investigation.

(1) The evaluation process, which requires about five hours and involves small groups of 8-12 men, includes a combination of objective and projective type personality tests, a diagnostic group session, evaluation and an assessee feedback period.

(2) As a result of his project, Dr. Mills has identified seven qualities which, although not all necessary in an overabundance, represent personality traits necessary for survival as a police officer.

(3) Although the Mills technique is not possible for use under current Army operations, it could be easily tailored for introduction into USAMPS operations in conjunction with the new one station training system in FY 77. The cost after initial study and consultant fees would be minimal in dollars and spaces. An exact amount estimate is impossible without further study into application of the Mills technique into MP screening requirements.

7. One of the foremost considerations in Army effort to set or renovate or modify police officer selection standards for prescreening techniques including psychological testing is the legal issue. Since 1972, the constraints placed on these efforts by the courts have been great. Review of the legal aspects of psychological testing as applied to law enforcement selection and other fields disclosed the following (Tab G):

a. On 24 Mar 72, Title VII of the Civil Rights Act of 1962 was amended by the Equal Employment Opportunities Act to achieve equality of employment opportunities.

b. Section 703(h) provides authority to give and act on results of any professionally developed ability test provided that such a test, its administration or action on results is not designated to discriminate. The employer will have the burden to show the reasonableness and necessity for the test.

c. If an examination disproportionately excludes a class of persons, the employer must satisfy the EEOC employment testing guidelines on examination validation.

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

d. Tests are defined as personal history/background investigations, specific education or work history requirements, scored interview, biographical information blanks, scored applications and interviewers rating scales.

e. Tests must be demonstrated by employers to be valid predictions of employee job performance. Validations can be achieved in one of two ways:

(1) Empirical validation showing strong correlation between test scores and job performance.

(2) Content construct validation involving an analysis of tests to be performed in the job in question on an examination built around the job analysis.

f. All tests reviewed including use of the Army ST score have the following deficiencies as screening technique for law enforcement personnel (except Tabs D & F).

(1) They measure and predict only scholastic performance.

(2) They are not built on a validated analysis of law enforcement job performance.

8. Alternatives. Based on the material reviewed and the findings surfaced in this study, the following alternatives are considered:

a. Continue utilizing exhibition of verified misconduct or undesirable behavior as the principal criteria for selection and/or retention of MP personnel. Inherent in this alternative is to continue utilizing the Army Classification Battery (ACB) test score of 100 as a minimum entry standard in the selection process.

(1) The current method of identifying substandard behavior is a subjective review of documents verifying undesirable behavior on the part of prospective or in-service law enforcement personnel which, to date, indicates that more than two percent of all personnel screened exhibit criminal or other undesirable behavior that should have precluded entry or retention in the law enforcement field.

(2) In addition to this system being subjective and expensive in terms of manhours, there is no capability for continuous surfacing of criminal or undesirable behavior.

(3) Lacking any specific criteria as to the degree of severity, the derogatory raw data are identified for additional review in the qualitative management process on a case-by-case basis.

(4) Once a man has been screened and accepted, there is no demand on the system to insure a file is automatically resurfaced for further evaluation on receipt of any derogatory information.

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

(5) Use of the ACB standard score 100 as selective criteria does not indicate characteristics favorable to predicting success in law enforcement or undesirable traits, nor predict with any acceptable validity academic success in law enforcement training. It is limited to prediction of success in a technical academic environment which recognizes basic knowledge of mathematics and English.

b. Find those characteristics which validly predict successful performance in the law enforcement field and to apply these standards in the selection process. Monitoring of the Cincinnati (Mills) and Chicago (Industrial Relations) studies which are endeavoring to identify valid selection characteristics in civilian law enforcement will provide a model for application in the selection of all service law enforcement personnel. More than simply providing yardstick criteria for evaluating overt criminal or undesirable behavior, these criteria utilized in the selective process will employ personnel managers to predict potential success or failure of an individual in the law enforcement field.

(1) Of the two studies, the Mills study appears to have the greater potential for detailed assessment of potential in law enforcement and is readily adaptable to the Army's proposed one-step training program. While simpler and less expensive, the Industrial Relations Center program is more adaptable for implementation at the recruiting stage of accessions to the law enforcement field. Adoption of the Mills study, even in a pilot program, would require refinement of enlistment contracts to stipulate that successful completion of the screening technique would be required prior to acceptance into the law enforcement field.

c. Find those characteristics which validly predict successful performance in the law enforcement field and to apply these standards in the selection process. It would now be possible to adopt the Mills and Industrial Relations programs utilizing their criteria without attempting to validate the same criteria in the service law enforcement environment. For reasons previously stated in alternative b, we need to develop criterion applicable to many law enforcement needs which may or may not be identical to civilian law enforcement. Further, time is also required to establish the predicted validity of the two studies in question.

9. Conclusions: While the primary thrust of this study was to examine the quality of military police as a factor adversely affecting law and order in the Army and the possibility of utilizing psychological testing in MP selection and screening the following ancillary conclusions have been surfaced.

a. Evidence has surfaced of immaturity in AIT graduates principally the result of the minimum age (18 yrs) established for entry into CMF 95B (Incl 1, Tab B). Civilian police studies indicate 21 years as the desirable minimum entry age to offset the immaturity factor. It is not practical to adopt this age as a minimum standard for military police entry because the preponderance of the Army accession age group is below 21. However, if the only negative criteria in the determination of an individual's long-range potential is a degree of immaturity correlated strictly to age, this temporary inadequacy can be largely corrected by improvement in quantity and quality of training and supervision.

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

b. At the time this study was being conducted the existing evaluation and screening program at USAMPS was inadequate because it provided only a superficial basis for identifying extreme cases of unsuitability and traditionally placed quantity output ahead of quality output (page 3, Incl 1, Tab B). While ongoing improvements in this area have been undertaken at USAMPS it is still believed a review and analysis of present screening techniques is needed to insure that present methods use quality as a key factor pending adoption of a more adequate psychological testing technique.

c. The uniformity and quality of military police military community interface is inadequate (Incl 1, Tab A) and can only be overcome by the development and promulgation of a broad base military police community relations program Army wide.

d. The present time constraints on the Basic Law Enforcement Training Program (seven weeks) while providing minimum preparation for military police performance in a combat support role also inadequately prepares the individual for performance in his broad base law enforcement role as indicated by military police supervisors in the field (page 5, Incl 2, Tab B). The adoption of any psychological screening technique will not overcome this training deficiency nor will subsequent in-service training serve as an adequate substitute. It would appear that a review of the adequacy of the seven week training program should be accomplished.

e. Selection standards currently utilized in the selection of military police are inadequately defined (Tab B). Existing standards must be reviewed and further developed before criteria for the prediction of success in a military law enforcement role can be identified and used to select success oriented law enforcement personnel for the Army.

10. Recommendations: It is recommended adoption of alternative b with implementation as follows:

a. Initiate study to identify, NLT end of 3d Qtr, FY 75, valid criteria for predicting success in service law enforcement.

b. Monitor progress of Mills and Industrial Relations Center studies through 3d Qtr, FY 75.

c. Based on a and b, be prepared to implement the pilot program utilizing techniques of the two which have proven more valid and easily adaptable.

d. That a review and analysis of the present screening techniques at USAMPS be conducted to insure that quality is the key consideration for screening out students.

e. That an Army-wide military police community relations program be developed and promulgated.

DAPE-HRE

SUBJECT: Recommendations 3F, Crimes of Violence

f. That the adequacy of the seven week basic law enforcement course at USAMPS be reviewed.

g. That existing standards for the selection of military police be reviewed and further developed toward the prediction of success in military law enforcement.

7 Incl
as

GLEN A. HILL
Colonel, MPC
Chief, Law Enforcement Division

A

DISPOSITION FORM

For use of this form, see AR 340-15; the proponent agency is The Adjutant General's Office.

REFERENCE OR OFFICE SYMBOL

DAPE-HRL

SUBJECT

, Recommendation 3f, Crimes of Violence

TO

DAPM

FROM

DAPE

DATE

11 JUL 1973
LTC Day/76912

CMT 1

1. One of the crime prevention recommendations approved by the Vice Chief of Staff on 5 September 1972 pertains to a study of the quality of military police as a factor affecting law and order. An initial study effort concluded that the scope of the problem should be expanded and that a broadly based working group be formed under the guidance of OTPMG. The 18 April 1973 status report to VCSA on this subject is at Incl 1.

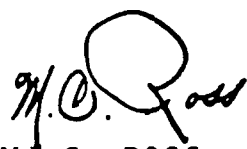
2. Request OTPMG form a working group comprised, as a minimum, of representatives from ODCSPER, MILPERCEN, OCRD, TRADOC and the US Army Military Police School, to:

"Determine the extent to which the quality of military police is a factor adversely affecting law and order in the Army and then, if warranted, the utility of using psychological tests in the MP selection and screening process."

3. Request the findings of the working group be provided ODCSPER, ATTN: DAPE-HRL, by 30 June 1974.

FOR THE DEPUTY CHIEF OF STAFF FOR PERSONNEL:

1 Incl
as


M. C. ROSS
Brigadier General, GS
Director of Human Resources
Development

RECOMMENDATION 3f

Determine, in coordination with the Special Assistant for Training OCSA, the extent to which the quality of military police is a factor adversely affecting law and order in the Army and then, if warranted, the utility of using psychological tests in the MP selection and screening process.

NOT YET COMPLETED. The study of the problem thus far has led to the determination that:

- a. There is insufficient information presently available to identify which characteristics in a policeman's makeup are associated with quality performance. Factors such as intelligence, even temperament, diligence, courtesy, self confidence, bravery and fairness are among the police characteristics frequently admired, yet, little is known scientifically about how these and other personality characteristics interact with forces which drive law and order. While acknowledging that some actions by a few military policemen may have caused the occurrence of incidents which might otherwise not have occurred, and these actions may have led to a degradation of the MP as viewed by some troops, quantification of these incidents has not been made.
- b. The discriminating utility of psychological tests will not help at this time in detecting defects in quality, even if it could be established precisely what is meant by "quality".

It is concluded that the effort should be continued, and the scope expanded to include:

- a. Determination of the meaning of "quality" as applied to military police personnel.
- b. Determination of the impact of quality personnel in law and order in the Army.
- c. Further review of the utility of psychological tests in the MP selection/screening process.

The effort should be directed by an ad-hoc working group chaired by OIPMG with representation from ODCSPER; MILPERCEN; CONARC; USAMPS; and OCRD. The time-frame should be extended to permit gathering of data and, as appropriate, the administering of tests. The estimated completion date is 3d quarter, FY 74.

THE QUALITY OF MILITARY POLICEMEN

FACTOR AFFECTING LAW AND ORDER

1. Introduction.

a. The quality of military policemen is a rather undefinable issue. There seems to be a fairly uniform understanding of what is meant by "quality;" however, when it comes to the point of defining what is a quality military policeman there is less uniformity. We can describe those attributes or characteristics which we think an MP should possess, and design our selection process accordingly, but of course there is no assurance that the result will be a "quality" MP. The question becomes even more complex when we wonder if the "quality" impacts adversely on law and order. The immediate emotional reaction is that it does — a poor military policeman means poor enforcement of law and order, but documentation to support this reaction is lacking.

b. Army Regulation 611-201, Enlisted Military Occupational Specialties, defines the physical, mental and special requirements for a military policeman. Aside from the physical requirements and some portions of the special requirements which are largely mechanical, requirements that can be termed "quality" fall into two general categories: undesirable behavior and innate qualities.

c. Undesirable behavior is the easier of the two to deal with. At the risk of over-generalization, it is a relatively simple matter to assess an individual's background of undesirable behavior. Granted that there are instances wherein individuals lie about or conceal their criminal background, or the Army fails to find out about an individual's behavior in a timely fashion, it remains fairly simple to make a quality determination of these individuals. Demonstrated undesirable behavior in the form of, say, arrests, permits a quick and easy assessment of the quality of a candidate to be a military policeman.

d. The potential MP is subjected to the same testing as any soldier; the Armed Forces Qualification Test (AFQT) and the Army Classification Battery (ACB) are the only "tests" which the soldier ordinarily takes. These tests are designed to measure aptitude and trainability, rather than innate qualities such as emotional control and initiative.

e. In the absence of a firm workable definition addressing a quality military policeman, this review is limited to factual evidence of military police exhibiting undesirable behavior, supervisor's opinions based on perceived job performance requirements, and community attitudes toward military police.

f. This section of the study will review and evaluate three areas impacting on the question of the quality of military police. The first area

is an opinion survey on community attitudes. Real or imagined, if the community being supported harbors negative attitudes toward its law enforcement body then a problem exists. The second area examined is a survey of military police supervisors' attitudes toward the quality of military police in CONUS. Although the reasons for these attitudes may be varied (i.e. training inadequacies, substandard individuals, prescreening shortcomings), the presence of any negative attitude reveals a confidence problem. The last area is a review of factual data (records and file reviews) to surface evidence of undesirable behavior. Conclusions drawn from the three reviews above will be shown in the last part of this section.

2. Opinion Surveys.

a. Opinions may be real based on factual data or perceived based on feelings, ideas, not necessarily factual in origin. Whether real or perceived, when a person has an opinion as to the quality and effectiveness of the military police, it will have an impact on any community law enforcement program. There are two opinion surveys done between 1972 and 1973 which impact on the question of quality and military police. The first is an August 1972 survey prepared by the US Army Military Personnel Center, Personnel Management Development Directorate, OPD-126, as of 31 Aug 72. The second is a survey team observation done by the Task Analysis/Training Programs Branch, USAMPS, between 12 Jun and 17 Jul 73. Neither of these surveys represents an in-depth product but do yield results which impact on perceived ideas of quality.

b. Each quarter the US Army Military Personnel Center conducts surveys of Army personnel to sample opinions, detect trends and measure effects of DA program. The DAPC-PMP Report No 17-73E represents a survey estimate of opinions on military police (MP) courtesy, image and professionalism conducted in Oct 73. The survey is considered statistically reliable although some terminology is questionable as to clarity. Nevertheless, it represents the thinking of Army members as best they understood the meaning of both the questions and answers. General results were as follows:

(1) More than 14 percent of the enlisted men in the Army surveyed (all grades) regard a military policeman as an "enemy."

(2) More than 45 percent of the Army officers and more than 50 percent of enlisted personnel surveyed do not feel military police are effective in a law enforcement/crime prevention role.

(3) More than 40 percent of the officers and more than 60 percent of the enlisted men surveyed regard the threat of violent crime as a menace to their personal safety on post.

(4) The image of the MP on qualities of helpfulness and courtesy leaves room for considerable improvement.

A more detailed analysis of the survey results is at Incl 1. The survey base for 1972 survey was 20,336 (N=20,336). Whether N responses were based on factual data, personal experiences or misunderstanding of role requirements, percentages reflecting a negative report with military police indicate the existence of a quality problem.

c. During the period 12 Jun to 17 Jul 73, officers of the Task Analysis/ Training Programs Branch, USAMPS, visited 33 units at 17 installations in CONUS in conjunction with systems engineering of the Basic Law Enforcement Course. The purpose of the trip was three-fold:

(1) To administer survey questionnaires concerning basic military police training requirements.

(2) To brief key personnel on the systems engineering program.

(3) To solicit comments, ideas and suggestions from knowledgeable individuals in the field on military police training (i.e. provost marshal, unit commanders, senior supervisors, etc). A comparison of the notes among the five officers comprising the survey teams developed the following problem areas:

(a) A significant number of MOS 95B20 and 95C20 replacements graduating from USAMPS demonstrate they are too immature to function effectively.

(b) Inadequate selection standards, haphazard screening and apparent low elimination criteria for students who are poor achievers during USAMPS training have resulted in a significant number of incapable military policemen and correctional specialists in the field.

(c) Military police graduates of the Basic Law Enforcement Course are not adequately trained in key functional police tasks.

(d) A large percentage of correctional specialists are not motivated to perform corrections duties.

This survey, while less formal than the previous one, shows an indication that provost marshals, unit commanders and senior supervisors feel that some of the new assessments into the Career Management Field 95 fall short of meeting the quality standards expected. A more detailed discussion of the problems cited above is shown at Incl 2.

3. The Qualitative Standards Management Program.

a. The Qualitative Standards Management Program was initiated in Sep 72 by Office of The Provost Marshal General. The object of the study was to determine if there were substandard individuals serving as military police who should never have been selected or retrained. The study base at the time of this review was 15,227 (N=15,227). The period of the study included

in this review is Sep 72 to Mar 73. The program included a 100 percent files check of all enlisted personnel holding primary military occupational specialty (PMOS) 95B. These files checks were conducted with the assistance and cooperation of the US Army Criminal Investigation Division Command, Crime Records Directorate at Fort Holabird. Based on the results of the check, recommendations would be submitted to the Military Personnel Center to originate directed reclassification of personnel identified as unsuitable. In the pilot study (first phase) from Sep to Oct 72, 4,500 files were selected at random for the check. Results revealed 428 derogatory files (9.5 percent). Minor juvenile or traffic offenses made up half of the derogatory files (209) while files containing disqualifying derogatory information made up the other half (219).

b. From the initial results it was clear that there were some lesser quality personnel serving as military police personnel. Thus far, approximately two percent of all files checked to date (N) were recommended for reclassification. Reasons for this problem which created this problem have been identified as follows:

- (1) Stress on quantity, instead of quality, during the Vietnam era.
- (2) Lack of quality control over lateral entry procedures.
- (3) Failure to reclassify individuals after demonstrated misconduct.

in this review is Sep 72 to Mar 73. The program included a 100 percent files check of all enlisted personnel holding primary military occupational specialty (PMOS) 95B. These files checks were conducted with the assistance and cooperation of the US Army Criminal Investigation Division Command, Crime Records Directorate at Fort Holabird. Based on the results of the check, recommendations would be submitted to the Military Personnel Center to originate directed reclassification of personnel identified as unsuitable. In the pilot study (first phase) from Sep to Oct 72, 4,500 files were selected at random for the check. Results revealed 428 derogatory files (9.5 percent). Minor juvenile or traffic offenses made up half of the derogatory files (209) while files containing disqualifying derogatory information made up the other half (219).

b. From the initial results it was clear that there were some lesser quality personnel serving as military police personnel. Thus far, approximately two percent of all files checked to date (N) were recommended for reclassification. Reasons for this problem which created this problem have been identified as follows:

- (1) Stress on quantity, instead of quality, during the Vietnam era.
- (2) Lack of quality control or lateral entry procedures.
- (3) Failure to reclassify individuals after demonstrated misconduct.

c. While the program is an ongoing one, to date, it has provided a factual data base on which to confirm or deny the presence of lesser quality individuals in the Military Police Corps. A more detailed explanation of the OTPMG Qualitative Standards Management Program developed by the Project Officer, LTC Edward D. Lockwood, Plans and Law Enforcement Division, OTPMG, is shown at Incl 3.

4. Conclusions.

a. There exists factual data on which to conclude that substandard individuals have been surfaced in the Military Police Corps during the past 18 months.

b. The number of lesser quality personnel identified for reclassification has been less than two percent of the Corps.

c. Evidence exists that in the opinion (real or perceived) of officers and enlisted men of the military community, the military police are ineffective in law enforcement/physical security roles.

d. Evidence exists that provost marshals, MP unit commanders and senior MP supervisors in CONUS feel a significant number of military police new assessments lack enough maturity, adequate training and motivation to be effective.

e. There are indications that selection standards, screening techniques and elimination procedures for lesser quality individuals are inadequate.

f. There exists no valid definition of what is quality military policemen.

g. That any major entry standards revision could be developed after 2d Qtr, FY 75 without legal complications.

QUESTION 1

Compare helpfulness of military and civilian police in your experience:

- A. Military police are more helpful.
- B. Civil police are more helpful.
- C. About the same.

RESPONSES

	<u>06-04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	37.4	32.7	33.9	22.7	19.1	19.9
B	12.6	17.0	15.6	18.1	24.0	22.6
C	50.0	50.3	50.5	59.2	56.9	57.5
* * * * *						

QUESTION 2

Compare courtesy rendered by military and civilian police in your experience:

- A. Military police are more courteous.
- B. Civil police are more courteous.
- C. About the same.

RESPONSES

	<u>06-04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	38.4	34.4	35.1	21.6	19.8	20.2
B	15.2	21.8	19.7	25.3	30.6	29.3
C	46.4	43.8	45.2	53.1	49.6	50.5

ANALYSIS

The two questions above are analyzed as a unit, since they are similar in purpose and content, and were used in this OPO survey to verify data collected in the Nov 71 survey. In Nov 71, responses indicated notable deficiencies in the helpfulness and courtesy of military police as seen by responders.

The two questions were widely separated in the survey itself. One appeared as Question 11, the other as Question 55. Thus the close relationship of the scale of responses reinforces the validity of the data.

Incl 1 To TABB

Less than 50 percent of all responders distinguished a difference between MP's and civil police in helpfulness. Among senior officers, MP's are seen more helpful than civilian police by a ratio of 3-1; MP's are viewed more helpful by a ratio of more than 2-1 among all officers. Among senior (E6-E9) enlisted personnel, MP's are also seen more helpful, but the ratio is much closer to 1-1. Civil police are seen as more helpful than MP's by the majority of junior enlisted personnel, but the ratio (about 4-5) is still close.

Although the scale of responses to the question on courtesy resembles the scale on the question on helpfulness, there are significant differences. First, a larger number of all responders distinguished between MP's and civil police. At least 50 percent of all responders (except senior enlisted men) picked either MP's or civil police as more courteous. Although MP's are seen as more courteous by a comfortable majority of officers, the ratio is only 2 1/2-1 among senior officers, and substantially less than 2-1 among all officers. Among all enlisted men, civil police are viewed as more courteous than MP's by a very significant majority: 5-4 among senior enlisted men, 3-2 among junior enlisted men.

Thus, data collected in Nov 71 has been verified and clarified. The image of the MP, compared to civil police, is very good on a scale of helpfulness among officers, but, considering the emphasis placed on training in military courtesy, only fair on a scale of courtesy. Among enlisted men, improvements in both helpfulness and courtesy are needed, but substantially greater stress should be placed on upgrading the image of the military police with regard to courtesy.

QUESTION 3

I see a military policeman as

- A. A policeman.
- B. A fellow soldier.
- C. A friend.
- D. An enemy.

RESPONSES

	<u>06-04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	52.2	60.8	57.0	44.6	45.0	45.0
B	36.1	28.9	32.0	38.6	31.7	33.3
C	7.5	4.0	5.4	7.6	7.5	7.5
D	4.2	6.3	5.5	9.2	15.8	14.2

ANALYSIS

A "Policeman" is unqualifiedly an authority figure. Most officers, and nearly a majority of enlisted men see the MP in this role.

A "Fellow Soldier" was designed to measure the degree of identification of responders with the MP. Unfortunately, it is clear now that the word "Fellow" was not interpreted in the same way by all responders. I believe this discrepancy invalidates most of the data collected on this question.

A "Friend" was not selected by a statistically significant number of responders.

An "Enemy" is a clear and incontrovertible term. Data collected on this response remains valid. The number of officers selecting "An Enemy" is higher than anticipated, but not high enough to cause concern. However the percent of enlisted personnel who view an MP as "An Enemy" is very disturbing, and defies analysis.

The figure of "more than 14 percent" of all enlisted men who consider an MP as an enemy can be useful in justification of recommendations for improved MP personnel standards, better training, or enhanced professionalism.

QUESTION 4

In your opinion, MP's are

- A. A pretty good police force.
- B. Good at some things, but not very effective at crime prevention and law enforcement
- C. Not very good at all.

RESPONSES

	<u>06.04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	57.3	52.7	54.2	53.2	46.5	48.0
B	36.6	40.4	39.3	38.7	38.1	38.3
C	6.1	6.9	6.6	8.1	15.4	13.7

ANALYSIS

The results of this question speak for themselves. Only slightly more than half of the officers, and less than half of the enlisted men see military police as "a pretty good police force." This phrase was used to stress the law enforcement aspects of MP operations, and the disappointing percentage who picked this response clearly points to a deficiency in the law enforcement image of military police.

This conclusion is reinforced by the substantial percentage who specifically identified a usefulness of military police (presumably in tactical operations, traffic control, security, etc.), but found the MP's wanting in law enforcement/ crime prevention.

Although unfavorable, the percentage of personnel at all grades who selected "not very good at all" was predictable in view of responses to other questions.

This data clearly identifies a significant problem - the "publics" view of the capability & effectiveness of military police to enforce law and prevent crime is not at all satisfactory.

QUESTION 5

If you could change one aspect of military police operations in the Army; what would you change?

- A. I would recruit only older men.
- B. The appearance of MP's.
- C. Training of MP's.
- D. Police Procedures
- E. Other

RESPONSES

	<u>06-04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	22.7	15.8	19.2	30.0	11.2	15.6
B	5.2	6.3	5.8	6.6	4.9	5.3
C	29.4	30.1	29.7	25.4	18.7	20.3
D	17.9	22.8	20.7	19.7	31.3	28.6
E	24.8	25.0	24.6	18.3	33.9	30.2

ANALYSIS

The fairly low percentage of respondents who selected "recruit only older men" is instructive. Only among senior enlisted men (30 percent) did the percent of selection approach a high figure. Thus, it appears that age of military police is not viewed as a major problem.

Junior enlisted men selected "police procedures" as the single most important factor requiring change. This may reflect dissatisfaction in general with authority, or with police in particular among a sizeable segment of young men.

The most significant finding is the high percentage of knowledgeable (officers and senior EM) who identified MP training as requiring improvement.

Overall, this was not a good survey question. Distractors did not provide sufficient flexibility, and this is demonstrated in the high percentage who selected "other" as their response. Except for adding emphasis to the ever-current need to achieve optimum training standards, I see no use this agency can make of this data.

QUESTION 6

On your installation, how do you feel about your safety or the safety of your dependents?

- A. Perfectly safe at all times.
- B. Feel safe during the day & unsafe in the evening.
- C. Violent crime is a serious threat to safety, but I take no precautions.
- D. Violent crime is a serious threat to safety and I take precautions accordingly.

RESPONSES

	<u>06-04</u>	<u>03-01</u>	<u>Officers</u>	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	60.7	56.5	57.7	38.0	38.8	38.7
B	21.5	28.5	25.8	32.8	28.5	29.5
C	1.1	1.8	1.5	1.9	3.9	3.4
D	16.7	13.2	15.0	27.3	28.8	28.4

ANALYSIS

The question was intended to clarify data collected in the Nov 71 survey. At that time, 25 percent of the senior officer respondents, and 75 percent of the junior EM indicated they felt unsafe in their quarters, or on the street, at night.

Response Alfa is considered desirable. Responses Bravo and Delta are degrees of fear and threat of harm. Response Charlie is identical to Response Delta, and was used to off-set the danger some respondents would be ashamed to admit, even anonymously, that they felt threatened. Responses Charlie and Delta are considered one in this analysis.

Among senior officers, nearly 40 percent express concern for their safety. Slightly more than half isolate the threat to the hours of darkness; the remainder feel violent crime is a threat at any time. This reflects a substantial deterioration since Nov 71.

Concern increases proportionate to decrease in rank. More junior officers express their awareness of danger than senior officers, but the differences throughout all commissioned grades are not significant.

The major increase occurs among enlisted men. With remarkable consistency through all ranks, less than 40 percent of the enlisted respondents feel secure on post.

For the senior grades, this represents a deterioration since Nov 71, while it reflects a slight improvement among the junior grades.

There are several conclusions available. First the threat is viewed differently by officers and EM, with little differentiation by grade. There appears to be a remarkable similiarity in the pattern of views of privates and senior sergeants, and of lieutenants and colonels. Second, officers, especially senior officers, have become more conscious of a threat to their safety since Nov 71. Senior enlisted men have also become more aware of a threat, but among junior enlisted men, consciousness of a personal menace has diminished slightly. The latter phenomenon may be due to personnel becoming more accustomed to a dangerous environment.

Response to this question clearly demonstrates that the Army is not providing an environment of safety to its personnel, and that the awareness of this condition among Army personnel is evident. Data can be cited as further justification for many types of recommendations to improve law enforcement, crime prevention, personnel, professionalism, etc.

QUESTION 7 (enlisted only)

Would you like to become a military policeman?

A. Yes.

B. No.

RESPONSES

	<u>E9-E6</u>	<u>E5-E1</u>	<u>Enlisted</u>
A	19.9	28.0	26.1
B	80.1	72.0	73.9

ANALYSIS

Question submitted for only one purpose: to measure potential for in-service recruiting (lateral entry) to fill law enforcement shortages.

Response clearly indicates a fertile field for in-service recruiting among junior EM. I plan to request OPO to break-out the response by each enlisted grade in their final report.

MP PROBLEMS

I. Problem: A significant number of 95B20 and 95C20 replacements graduating from USAMPS demonstrate they are too immature to function effectively.

a. Problem Discussion:

(1) There is a consensus among commanders and provost marshals visited by survey teams that many of the replacements lack maturity necessary to cope with their assigned tasks.

(2) One of the most detrimental by-products perceived in immature replacements is an inability to cope with stress. Low stress tolerance adversely affects judgment and, in some cases, causes overreaction to routine situations. The most classic example of overreaction is when a young MP takes a traffic violator's hostile or critical attitude personally and turns a routine citation into a heated confrontation. The same type of overreaction occurs in correctional facilities between correctional specialists and prisoners.

(3) Lack of confidence is another deficiency observed in replacements which is attributed in varying degrees to immaturity. Correctional specialists in worrisome numbers are hesitant to give prisoners orders. Frequently they are over solicitious to prisoners, rendering themselves virtually ineffective. A youthful appearance contributes to the lack of

confidence in both 95B20 and 95C20 personnel. A young face immediately puts the 95B at a disadvantage with an older offender and a 95C at a similar disadvantage with an older prisoner. This is overcome only by a demonstrated mature and professional manner on the part of the youthful 95B or 95C personnel.

b. Conclusions/Recommendations:

(1) Several persons interviewed felt the first step necessary to overcome the immaturity perceived in replacements is to develop a means to determine the maturity level of prospective military policemen and correctional specialists through some kind of psychological test or evaluation. Civil police departments in ever increasing numbers are employing this type psychological testing. An alternative, or additional action, would be to increase the minimum age requirements. Usefulness of this latter course is debatable. It seems reasonable that some 18 year olds would be more mature than other 20 year olds.

(2) Some of them felt the associated problem of low stress tolerance can, to a large extent, be corrected through the school curriculum. The capacity to cope with stress can be developed through more "hands on training" emphasizing performance under stress. The MP or CSP student should be confronted in the classroom with the hostility and pressure he may experience on the job. This allows the students to know what to expect

from those confronting them and to know what their own feelings or reactions may be and how to deal with these feelings.

(2) Repetitious training helps to develop confidence. The School currently employs "hands-on-training" but, unfortunately, the 7-week curriculum just doesn't allow the amount necessary. A person develops confidence through the successful repetition of an act. A training situation, once experienced, can be more easily transferred to a similar actual situation in the field.

II. Problem: Inadequate selection standards, haphazard screening and apparent low elimination criteria for students who are poor achievers during USAMPS training have resulted in a significant number of incapable military policemen and correctional specialists in the field.

a. Problem Discussion:

(1) The two most disagreeable elements among the selection standards, according to those commanders and provost marshals sampled, are the minimum age and insufficient educational development. They felt that a high school background or its demonstrated equivalent is essential to grasping the complexities of the military police or correctional specialist mission. Both must make unsupervised decisions and possess the ability to give written and oral reports with clarity.

(2) Several commanders reported having received replacements who had civil felony convictions on their record. The fact that these individuals slip through the system can in part be linked to over zealous recruiting. Apparently, there is too much emphasis on quotas.

(3) Many commanders and provost marshals felt that "duds" are not eliminated at USAMPS with the consistency necessary to maintain appropriate standards. They went on to suggest that the unwillingness to eliminate the incompetent and unfit during training may stem from the misconception that these individuals fail because of cadre failures and not because of their own inadequacies. Maybe the cadre (instructor) feels that there is a stigma attached to washing a man out which will reflect on his ability to teach.

b. Conclusions/Recommendations:

(1) The apparent solution to inadequacies in selection standards is to set standards at an appropriate level and then insure the recruiting system is not allowed to deviate from these standards. The minimum age is arbitrary and contingent on several considerations, but the need to require a high school background or its demonstrated equivalency for acceptance to 95B20 and 95C20 MOS training is unquestionable. The fact that not all high school graduates are equally qualified must also be considered. A test designed to ascertain an individual's ability to perform military police or correctional specialist duties may be appropriate. Results of the 95B/95C

systems engineering field survey can be used to document this need as evidenced, first of all, by opinions of provost marshals and commanders surveyed and by comments from lower and middle level supervisors on the weaknesses of 95B20 and 95C20 incumbents. Several of those weaknesses are discussed as functional area weaknesses in paragraph III.

(2) Screening procedures will be improved only through emphasizing strict quality control. Priority must be given to recruiting individuals who can get the job done rather than merely filling quotas. "Unit of choice programs" must be conducted on a quality rather than quantity basis.

(3) Every effort must be made to eliminate those who cannot measure up at the earliest possible point in their training at USAMPS. Special evaluation techniques need to be developed to achieve this goal.

III. Problem: Military police graduates of the Basic Law Enforcement Course are not adequately training in key functional police tasks.

a. Problem Discussion:

(1) Report writing is, according to field feedback, the major functional deficiency observed in 95B20 personnel. Many military policemen cannot present the information required in a police report effectively. For some

military policemen this deficiency can be traced to poor educational background. For the others it is a problem of concept. The total problem is amplified by inconsistencies in report writing requirements at various posts throughout the Army.

(2) Operation of TCP's is another major weakness observed in 95B20's. They are normally proficient in technique but lack the confidence necessary to control a busy intersection.

(3) Several of the provost marshals interviewed expressed concern that military policemen are inadequately training with the .45 caliber pistol. A significant number have no confidence in the weapon and, in their estimation, would not be capable of reacting quickly and effectively in a time of emergency.

(4) Some military police replacements are not driver qualified. Although the number of these individuals is comparatively small, it is inconceivable to commanders and provost marshals that anyone could graduate from USAMPS without being capable of operating a vehicle, even if it is only a 1/4 ton. Without driver qualification, their usefulness to the unit is very limited until they can be trained. Most commanders and provost marshals recognize our sedan problem and realize this situation is beyond our control.

(5) Other frequently reported functional weaknesses are investigative skills (to include traffic), a knowledge of laws important to the police mission and use of military police forms in general.

b. Conclusions/Recommendations:

(1) Performance deficiencies in the functional areas discussed above can be corrected by additional or redirected training. The time for this additional or redirected training can be provided by a greater emphasis on police tasks (advocated by every PM interviewed and a great majority of the commanders) and lengthening the Basic Law Enforcement Course. Specific recommendations on what should be taught and length of time necessary to teach each task will be products of the systems engineering of the basic course.

(2) Correction of report writing deficiencies appears to be contingent on more and/or better training. Training provided must include a basic understanding of what is needed to have a complete report, as well as how to fill in the blanks on a given form.

(3) Traffic control efficiency can be upgraded through additional training. Since the root of this problem is confidence, training should be designed to develop confidence as discussed in paragraph I.

(4) It appears that weapons qualification for military policemen should encompass more than just mechanically firing at stationary targets. Weapons training should be designed to develop overall confidence in the weapon and then the training should include firing under circumstances in which parallel

situations may actually exist in confrontations with offenders. Several provost marshals and commanders suggested an FBI "type" qualification course.

c. Commanders and provost marshals feel the solution to the remaining deficiencies presented above is contingent on continued application of the dual theme of this paragraph: increased training time spent primarily in "hands on training."

IV. Problem: A large percentage of correctional specialists are not motivated to perform corrections duties.

a. Problem Discussion: A large number of incumbent correctional specialists were selected randomly from student companies at USAMPS without having volunteered for the corrections field. Many of the "chosen" were primarily interested in 95B duty and openly reject corrections. This is manifested in a generally poor attitude affecting courtesy and appearance and their handling of prisoners.

b. Conclusions/Recommendations: The practice of arbitrarily assigning students in the basic course who have not specifically enlisted for 95C is counterproductive. The answer, unfortunately, is one that the recruiting command must develop.

**QUALITATIVE STANDARDS MANAGEMENT
(RECLASSIFICATION PROJECT)**

THE REQUIREMENT THAT LAW ENFORCEMENT PERSONNEL POSSESS RECORDS FREE OF MISCONDUCT - PARTICULARLY OF A CRIMINAL NATURE - IS LONG STANDING AND VIRTUALLY ACCEPTED AS AN INCONTRAVERTIBLE FACT. THE PUBLIC MUST HAVE THE CONFIDENCE THAT ITS LAW ENFORCEMENT PERSONNEL CAN BE TRUSTED, AND THAT RIGHT APPLIES EQUALLY TO A MILITARY "PUBLIC" AS TO A CIVILIAN "PUBLIC." IN ADDITION, THE CREDIBILITY OF A POLICE OFFICER MUST STAND ON ITS OWN, BOTH IN THE FACE OF THE COMMUNITY SERVED, AND ESPECIALLY IN A COURT ROOM CONTEXT.

ABOUT A YEAR AND A HALF AGO, HQDA BECAME AWARE THAT THESE STANDARDS HAD BEEN PERMITTED TO SLIP IN THE SELECTION AND RETENTION OF MILITARY POLICE. THE FIRST SIGNS WERE REPETITIVE INSTANCES OF MISCONDUCT ON THE PART OF MILITARY POLICE PERSONNEL IN WHICH RESEARCH INDICATED THE OFFENDING MP SHOULD NEVER HAVE BEEN SELECTED OR RETAINED. AS A RESULT OF THESE SIGNS, IT WAS DETERMINED NECESSARY AND ADVISABLE TO CONDUCT A FULL scale STUDY OF THE PROBLEM.

THE STEPS TAKEN IN THE CONDUCT OF THE STUDY WERE AS FOLLOWS:

1. A COMPLETE EMTR (ENLISTED MASTER TAPE) PRINT-OUT ON ALL ENLISTED PERSONNEL HOLDING PMOS 95B WAS REQUESTED FROM (THEN) PERSINSCOM. THIS WAS COMPLETED IN SEPTEMBER 1972, AND A FULL TAPE, PLUS PUNCH CARDS, WAS PROVIDED TO OTPMG.

2. WITH THE CONCURRENCE OF THE COMMANDER, USACIDC, 4,500 PUNCH CARDS, SELECTED AT RANDOM, WERE PROVIDED TO CRIME RECORDS DIRECTORATE AT FORT HOLABIRD FOR A RECORDS CHECK. ACTION WAS COMPLETED IN OCTOBER

3. RESULTS WERE ALARMING. ALMOST TEN PERCENT OF THE PERSONNEL POSSESSED DEROGATORY FILES, AND SLIGHTLY MORE THAN HALF - ABOUT FIVE PERCENT - CONTAINED DEROGATORY INFORMATION OF A NATURE SUFFICIENT TO DISQUALIFY THEM FROM LAW ENFORCEMENT SERVICE. This data is depicted AT INCLOSURE 1.

IN NOVEMBER, 1972, THE DIRECTOR, ENLISTED PERSONNEL DIRECTORATE (EPD) WAS BRIEFED ON THE FINDINGS OF THIS PRELIMINARY STUDY, USING THE DATA I HAVE JUST DESCRIBED. DURING THAT BRIEFING, IT WAS CONCLUDED THAT OTPMG, WITH THE ASSISTANCE AND COOPERATION OF USACIDC-CRD, WOULD CONDUCT A 100 PERCENT FILES CHECK OF ALL ENLISTED PERSONNEL HOLDING PMOS 95B. BASED ON THE RESULTS OF THAT CHECK, RECOMMENDATION WOULD BE SUBMITTED TO EPD (NOW PART OF MILPERCEN) TO ORIGINATE DIRECTED RECLASSIFICATION OF PERSONNEL IDENTIFIED AS UNSUITABLE.

AS A LONG RANGE OBJECTIVE, IT WAS DETERMINED A MAJOR EFFORT WOULD BE MADE TO REWRITE STANDARDS AND PROCEDURES TO PRECLUDE RECURRENCE OF THIS SITUATION.

IN ADDITION, IT WAS CONSIDERED NECESSARY AND APPROPRIATE TO PROCURE A WRITTEN TJAG OPINION REGARDING THE LEGAL SUFFICIENCY OF THE PROCEDURES ESTABLISHED FOR THIS DIRECTED RECLASSIFICATION PROJECT. ALTHOUGH THERE WAS SOME DELAY, TJAG RULED THE PROPOSED ACTION WAS LEGALLY AND ADMINISTRATIVELY SUFFICIENT ON 3 MARCH 1973.

CURRENT EFFORT, THEN, DATES FROM THE BEGINNING OF APRIL 1973. SINCE THAT TIME, MORE THAN 15,000 FILES HAVE BEEN REVIEWED. APPROXIMATELY 5600 FILES REMAIN TO BE REVIEWED. THE RESULTS, IN TERMS OF NUMBERS,

MORE IMPORTANT THAN THE NUMERICAL RESULTS, HAVE BEEN THE TANGENTIAL BENEFITS.

1. FIRST, RECOGNITION OF THE PROBLEM WHICH CAUSED THE CONDITION TO EXIST IN THE FIRST PLACE:

A. STRESS ON QUANTITY, INSTEAD OF QUALITY, DURING THE VIETNAM ERA.

B. LATERAL ENTRY PROCEDURES, ESPECIALLY IN TIMES OF ACUTE LOCALIZED SHORTAGES.

C. FAILURE TO RECLASSIFY INDIVIDUALS AFTER DEMONSTRATED MISCONDUCT. THERE HAVE GENERALLY BEEN THREE REASONS FOR THIS FAILURE:

(1) EXISTENCE OF AN "OUT OF SIGHT-OUT OF MIND" MENTALITY WHICH HAS JUSTIFIED TRANSFER TO ANOTHER UNIT AS A SUBSTITUTE FOR DECISIVE RECLASSIFICATION ACTION.

(2) INSUFFICIENTLY PRECISE WORDING OF THE STANDARDS WHICH HAVE PERMITTED WIDE INTERPRETATION.

(3) RECLASSIFICATION BOARDS SIMPLY FAILING TO RECOGNIZE THE SPECIAL REQUIREMENTS OF MP'S.

2. SECOND, SENSITIZING MILPERCEN REGARDING THE PROBLEM. AS A RESULT, A SUBSTANTIAL NUMBER OF SIDE BENEFITS HAVE ACCRUED:

A. MILPERCEN HAS PROPOSED 46 RECLASSIFICATIONS BASED ON DATA AVAILABLE TO THAT ORGANIZATION AND NOT CONTAINED IN INVESTIGATIVE FILES. WE HAVE CONCURRED IN 36 OF THESE RECLASSIFICATIONS, AND THEY HAVE BEEN PUT INTO EFFECT.

B. MILPERCEN HAS ORIGINATED A REQUIREMENT THAT ALL PERSONNEL IN GRADE E6 OR ABOVE (THE CENTRALLY MANAGED GRADES) DESIRING

LATERAL ACCESSION INTO CMF 95 BE CLEARED THROUGH HQDA. THESE PROPOSED RECLASSIFICATIONS INTO THE LAW ENFORCEMENT FIELD HAVE GIVEN US AN OPPORTUNITY TO COMMENT ON THE SUITABILITY OF THE INDIVIDUALS INVOLVED, AND FOR THE FIRST TIME, HAS PROVIDED A REAL CONTROL ON LATERAL ENTRY. THE RESULTS OF THIS REVIEW THROUGH 1 APR 74 ARE DEPICTED FOR YOU AT INCLOSURE 3.

C. AT OUR RECOMMENDATION, MILPERCEN IS CONSIDERING EXPANDING THE LATERAL ENTRY CHECK TO INCLUDE ALL GRADES.

THE FOLLOWING STEPS REMAIN TO BE TAKEN. I MIGHT ADD THAT, AT THIS TIME, WE DO NOT VISUALIZE DA STAFF REORGANIZATION ALTERING THIS SCENARIO IN ANY WAY.

1. COMPLETE THE CURRENT SCREENING. DELAY HAS NOT BEEN ENTIRELY ACCIDENTAL. ONE OF THE MAIN REASONS IS THE NEED TO RESPOND TO REBUTTALS ON A TIMELY BASIS. AS A RESULT, WE HAVE NOT PERMITTED TOO MANY DIRECTED RECLASSIFICATIONS TO BE IN THE FIELD AT ANY ONE TIME, BECAUSE OF THE CLEAR DANGER THAT OUR OWN ADMINISTRATIVE SHORTCOMINGS WOULD RESULT IN OUR FALLING BEHIND IN RESPONDING TO REBUTTAL ACTIONS.

2. WE MUST GO BACK AND PICK UP ALL PERSONNEL WHO ENTERED THE CMF 95 AFTER 31 AUGUST 1972 (THE DATE OF THE EMTR TAPE), AND ABOUT 1,000 PERSONS WHOSE FILES HAVE BEEN DELAYED BECAUSE THEY WERE CHECKED OUT BY ANOTHER AGENCY, OR IN PROCESS WITH AN OPEN CASE.

3. IN ADDITION, WE MUST PROVIDE INFORMATION TO MAJOR COMMANDS ON CERTAIN INDIVIDUALS WHOSE FILES REVEAL DEROGATORY DATA OF A NATURE THAT IS EITHER SENSITIVE (THERE ARE ONLY A HANDFUL OF SUCH FILES) OR UPON WHICH HQDA IS NOT IN THE BEST POSITION TO RENDER A JUDGMENT.

I SUPPOSE YOU MIGHT DESCRIBE THE LATTER AS ALL PERSONNEL WHOSE RECORDS PUT THEM INTO THE "PROVERBIAL GRAY AREA."

4. FINALLY, WE WILL UNDERTAKE A REVISION OF STANDARDS AND PROCEDURES FOR ENTRANCE INTO ACGP 95, AND RECLASSIFICATION. A VERY LOGICAL QUESTION IS WHY WAIT? WHY NOT UNDERTAKE REVISION OF STANDARDS NOW OR SOONER? WHY WAIT UNTIL THE RECLASSIFICATION IS COMPLETED? THE ANSWER IS TIED TO THE FACT THAT A CHANGE IN STANDARDS WOULD SERVE TO INVALIDATE THE CURRENT RECLASSIFICATION EFFORTS.

A. THE CURRENT STANDARDS ARE DEPICTED AT INCLOSURE 4.

B. THE PROBLEM CONNECTED WITH THEM IS TWO-FOLD: AS WRITTEN, THEY LACK SPECIFICITY; IN ADDITION, THERE ARE NO CLEAR-CUT AND INCONTRAVERTIBLE RULES THAT MAKE THESE STANDARDS APPLICABLE TO AN INDIVIDUAL AFTER HE OR SHE HAS ENTERED THE LAW ENFORCEMENT FIELD.

C. WHAT WE PLAN TO PROPOSE IS BASICALLY THIS:

(1) FIRST, WE HOPE TO TAKE ALL STANDARDS OUT OF AR 611-201, AND PREPARE A SEPARATE CHAPTER IN AR 614-200 (PERSONNEL MANAGEMENT) ENTITLED THE "MANAGEMENT OF LAW ENFORCEMENT PERSONNEL."

(2) SECOND, WE HOPE TO MAKE THE LANGUAGE OF THE NEW STANDARDS MORE CLEAR AND PRECISE. WE WILL INSURE THAT RECLASSIFICATION OUT OF THE MOS IS TREATED AS THOROUGHLY AND IN AS MUCH DETAIL AS ENTRANCE INTO THE CMF, FOR EXAMPLE.

(3) THIRD, WE HOPE TO DEVELOP TWO TYPES OF STANDARDS. THE FIRST WILL BE MANDATORY, AND NOT WAIVERABLE. IN THE CASE OF ENTRANCE, IT WILL BE FINAL AND UNQUALIFIED. WAIVERS WILL NOT BE AUTHORIZED. IN THE CASE OF AN IN-SERVICE DETERMINATION, IT WILL TRIGGER AUTOMATIC RECLASSIFICATION. AN EXAMPLE OF THIS KIND OF STANDARD WILL BE CONVICTION OF A FELONY OFFENSE. THE SECOND TYPE OF STANDARD WILL BE DISCRETIONARY. IT

WILL SERVE AS A BAR TO ENTRANCE INTO ACGP 95 WITHOUT A WAIVER, BUT A WAIVER WILL BE POSSIBLE. WAIVER AUTHORITY WILL BE CLEARLY SPECIFIED, AND A PROCEDURE TO INSURE THAT A COPY OF THE WAIVER BE MADE A PART OF THE SOLDIER'S PERMANENT FILE WILL BE ESTABLISHED. IN CONNECTION WITH AN IN-SERVICE OCCURRENCE, IT WILL REQUIRE THE COMMANDER TO PROVIDE CLEAR JUSTIFICATION FOR RETENTION OF THE INDIVIDUAL IN THE CAREER FIELD: IN OTHER WORDS, IT WILL REQUIRE THE COMMANDER TO ACT IN SOME WAY. IN ADDITION, THE LANGUAGE WILL COVER IN MORE SPECIFIC TERMS SOME OF THE GENERAL AREAS WE HAVE FOUND TROUBLESOME. FOR EXAMPLE:

- (A) FINANCIAL IRRESPONSIBILITY.
- (B) ALCOHOLISM AND USE OF DRUGS.
- (C) REPETITIVE MISCONDUCT OF A MINOR NATURE.
- (D) ABUSE OF POSITION OR AUTHORITY.
- (E) MORAL TURPITUDE.

ON BALANCE, WE ARE CONVINCED THIS PROGRAM IS GOING TO DO A GREAT DEAL TO IMPROVE THE OVERALL CALIBER OF LAW ENFORCEMENT PERSONNEL IN THE ARMY, AND WILL PROVIDE COMMANDERS AND PROVOST MARSHALS A VERY USEFUL TOOL IN THE YEARS AHEAD.

RECLASSIFICATION PROJECT

TEST RESULTS

WORLD WIDE 95B STRENGTH AS OF 31 AUGUST 1972.....	21,617
RECORDS SCREENED AS OF 31 OCTOBER 1972	4,502
	(20.8 %)
DEROGATORY FILES IDENTIFIED	428
	(9.5 %)
FILES CONTAINING ONLY MINOR, JUVENILE OR TRAFFIC OFFENSES.....	209
FILES CONTAINING DISQUALIFYING DEROGATORY INFORMATION	219

RECLASSIFICATION REQUEST

STATUS AS OF 1 APRIL 1974

ENLISTED FILES CHECKED	15,227
ENLISTED FILES REMAINING	5,651

FILES CONTAINING NO SUBSTANTIVE DEROGATORY INFORMATION	11,722
FILES REQUIRING NATIONAL AGENCY CHECK	1,427
FILES PENDING ACTION	1,463
RECLASSIFICATIONS RECOMMENDED ON BASIS OF FILE CHECK	327

REBUTTALS SUBMITTED	57
REBUTTALS SUSTAINED	13

RECLASSIFICATION PROJECT
MILPERCEN ORIGINATED ACTIONS
(SINCE 1 OCT 73)

RECLASSIFICATIONS

PROPOSED46
CONFIRMED BY OTPMG36
REBUTTED SUCCESSFULLY 3

ENTRANCE CHECKS

CONDUCTED176
APPLICANT CONFIRMED107
APPLICANT REJECTED 19
NATIONAL AGENCY CHECK REQUESTED 28
INQUIRY STILL PENDING 22

CURRENT MILITARY POLICE STANDARDS

(PAST RECORD ONLY)

AR 611-201

NO RECORD OF:

- 1. SPECIAL OR GENERAL COURT MARTIAL CONVICTIONS.**
- 2. MORE THAN 30 DAYS LOST UNDER SECTION 972, 10 USC, APPENDIX 3, MCM 1068. ("BAD TIME")**
- 3. CIVILIAN CONVICTIONS OTHER THAN MINOR TRAFFIC OFFENSES.**
- 4. CIVIL TRAFFIC CONVICTIONS OR TRAFFIC OFFENSES, EITHER CIVIL OR MILITARY THAT WARRANT ASSESSMENT OF MORE THAN SIX TRAFFIC POINTS FOR A SINGLE OFFENSE OR MORE THAN A TOTAL OF 12 TRAFFIC POINTS FOR TWO OR MORE OFFENSES AS DEFINED IN THE POINT ASSESSMENT TABLE, AR 190-5, DURING THE 12 MONTH PERIOD IMMEDIATELY PRECEDING THE DATE OF ENLISTMENT OR REENLISTMENT IN THE MP CAREER FIELD. (INTERPRETED TO ALSO INCLUDE ACCUMULATION OF 12 TRAFFIC POINTS IN A SINGLE 12 MONTH PERIOD AFTER ENTRANCE INTO MP CAREER FIELD.)**
- 5. ENLISTMENT WAIVER GRANTED BY THE ADJUTANT GENERAL. (RELATIVELY MEANINGLESS).**
- 6. UNDESIREABLE BEHAVIOR EVIDENCED BY ANY RECORD, CIVILIAN OR MILITARY.**

A REVIEW OF PSYCHOLOGICAL TESTING IN POLICE SELECTION

1. Background. It is axiomatic in law enforcement that the best man be chosen to perform the required tasks. Failure to so choose will assure inadequate or incompetent performance. Part of the selection problem deals with existing inability to clearly define tasks and evaluate personnel. The law enforcement selection process has employed such screening elements as residence, education, height, physical agility, written general intelligence tests, oral interviews, training programs, probation and attempts at psychological testing. The latter has varied from paper and pencil, superficial mental health, motivational tests to extensive projection and interview techniques by psychiatrists. In order to answer the question concerning use of psychological testing in selection of US Army law enforcement personnel, it is necessary to know what is available for consideration. This review is not intended to be exhaustive but a representation of work done in this field over the last 14 years. It is divided in three major sections. The first section addresses police selection studies that deal with the issue of validity. The second section deals with nonvalidity studies. The third section summarizes the present state of the art and conclusions.

2. Validity Studies in Police Selection.

a. Validity studies described here deal with selection of police officers and examines the relationship between job performance predictors (i.e. test scores) and actual performance (i.e. supervisory ratings). For the most part, predictors have fallen into one of four categories:

(1) Psychological tests of aptitude, intelligence, interest, personality, etc.

(2) Biographical information (i.e. age, education, employment history, marital status, race, etc.).

(3) Situational tests.

(4) Civil service examinations.

Measures of job performance fall into four categories:

(1) Achievement in police academy settings.

(2) Termination of service.

(3) Supervisory ratings.

(4) Objective measures (i.e. number of awards, arrests, complaints, rank progression, etc.).

b. In 1950, Dubois and Watson conducted an early study on patrolman selection.¹ Two classes of recruits (N=129) were given the following 10 tests: Departmental Police Aptitude Test, Army General Classification Test, Cornell Word Form Test, Figure Matching Test, Bennett Mechanical Comprehensive Test, Minnesota Paper Form Board Test, Object Aperture Test, Strong Vocational Interest Blank, Rosenzweig Picture Frustration Test, and a Handwriting and Composition Test. At the completion of the eight-week academy, test predictors were related to four measures: academy grades, achievement test scores, marksmanship, and supervisors rating. The Police Aptitude Test and Army General Classification Test were consistently good predictors of academic performance. The Police Aptitude Test correlated .39 and .50 with academy for the first and second recruit classes respectively. The correlation between grades and achievement test scores was .47 for the two classes combined. The Army General Classification Test correlated .54 and .50 with academy grades for classes one and two respectively and .47 with achievement test scores for the combined classes. The six correlations were considered significant. Combination of the above two tests and either the Bennett or Paper Form Board yielded significant multiple correlation co-efficients of approximately .60 with both academy grade and achievement test scores. No predictor test was significantly correlated with the supervisory rating criterion. Cross-validation with a third class (N=36) showed a multiple correlation of .62 with final grades in the academy.

¹

DuBois, Philip H. and Watson, Robert I, The Selection of Patrolmen, Journal of Applied Psychology 1950, 34(2), 90-94.

The prime result of DuBois and Watson work was that certain tests can be used to predict later academy performance. During the same period, other studies attempted to apply existing test and methods to the selection of law enforcement personnel. Levy's 1960 report concedes that it is necessary to recognize the need to eliminate police candidates whose personality structures may make them poor risks for police service, but she failed to identify a methodology capable of elimination. She concluded that what is needed is identification of the interplay between personality components and the identifiable stresses of police work.² In her later study (1970) Levy examined 5,000 personnel files of police officers to develop a predictive model of tenure for new officers. Her model had predicted termination of recruits with some efficiency.³ Value of this study is doubtful as it has encountered substantial validity criticism in a 1970 study by Newmen.⁴ The City of Baltimore found a relationship between achievement on examinations and success in police training school expressed through use of

2

Ruty Levy, Report on Research Project, Occupational Selection of Police Officers (San Jose: Jan 4, 1960), pp 1-5 (mimeographed)

3

Ruty Levy, Investigation of a Method for Identification of the High Risk Police Applicant, Berkeley, CA, Institute for Local Self-Government 1970

4

Newman et al, Can High Risk Police Applicants be Identified? Berkeley, CA, Institute for Local Self-Government, Jul 1971, pp 59-89

psychological tests. Tests were the old Army AGCI, spelling tests and penmanship tests correlated with results of personal interview.⁵ The 1961 study of Blum, Goggin and Whitmore stated that for purposes of continuing police selection research the following test methods seem worthy of further study: The Rorschach Ink Blot Test, the Minnesota Multiphasic Personality Inventory, The F Test, the Cordell Test of Practical Judgment and the content analysis of an essay on reasons for wanting to be a policeman, and a life history, summary and observation in the test situation.⁶ They indicated other psychological tests which do not seem to merit further research at this time.⁷

c. Colarelli and Siegal (1964) administered the California Test of Mental Maturity, the all Port-Vernon-Lindzey study of values, the Edwards Personal Preference Schedule, and the Minnesota Multiphasic Personality Inventory to Kansas State Highway Patrol applicants. From all of this, a description of the average patrolman was drawn. Eight job performance variables (moving hazardous arrests, moving hazardous warnings, other arrests, services rendered, light correction, miles per contact with and without radar, and hours per arrest) were summarized for each patrolman to

⁵ Ronald A. LaCouture, Merits of Application of Psychological Testing, The Police Chief XXVII (Aug 60) pp 20-21

⁶ Richard H. Blum, William L. Goggin and Earl Whitmore, A Study of Deputy Sheriff Selection Procedures, Police (Reprint) (Nov-Dec 1961)

⁷ IBID

provide an opportunity measure which should have controlled the differences between population of the community served. Success or failure predictions were based on correlations between test scores and criteria measures. The control group (N=60) all took the same test battery. Supervisors who were unaware of the predictions rated the recruits at a later time. Results were reported in general terms. All but one of the predicted failures were rated as poor police quality, while all predicted successes were rated above average and higher. Despite the apparent success of this study, it cannot be classified better than questionable because the eight job performance missions were not validated. No specific correlations, test of significance or success or failure numbers were given.⁸

d. Dr. Aurelins Abbatiello, Director of Examination, City of Chicago (1969) reported a study using the Civil Service Examination as the predictor and a multiple criterion consisting of failure to qualify (FTQ) on the background examination and achievement in police academy measured by both grades and instructor ratings. A negative correlation of .17 was found between the Civil Service Examination score and FTQ. Significant correlations of .35 and .25 were found between Civil Service Examination scores and academy grades and Civil Service Examination scores and instructor ratings. Abbatiello concluded that the Civil Service Examination had predicted

⁸Colarelli, Nick J. and Siegal, Saul M., A Method of Police Selection. Journal of Criminal Law, Criminology and Police Science, 1964, 55 (2) 287-289

scholastic performance in the Police Academy.⁹ Use of the Civil Service Examination scores as a selection criteria for police officers is currently under fire in Chicago. Dr. Abbatiello feels today that this approach is not the answer. He feels the answer is a pure law enforcement job related test which will surface and place value on "Alley Awareness" (the feel for the ways of the street) while considering cultural differences.¹⁰ In a later study involving the Civil Service Examination, Spencer and Nichols (1971) reported a significant correlation between the Management Psychologist, Inc. (MPI) and FTQ (.24). A correlation of .18 was reported between the MPI and department performance rating. Again, the result was predicted scholastic performance only.¹¹

e. One of the better tests involving education and maturity done in the mid to late sixties was by Morman et al, 1966-67. They used a series of studies with the TAV Selection System. This test or system consists of statements concerning personal data, judgment, proverbs and sayings, preferences and adjective checklist, yielding 15 predictor scores. Results, reported in terms of multiple correlation, indicated a correlation between

⁹ A. Abbatiello, A study of police candidate selection. Paper presented at 78th annual convention of the American Psychological Assoc, Wash, DC, 1969

¹⁰ Personal Conversation, Dr. Abbatiello, Jan 74

¹¹ Spencer, Gilmore and Nichols, Robert A., A Study of Chicago Police Recruits: Validation of Selection Procedures, Police Chief, 1971 38(6), 50-55

age, amount of experience, education and test predictors of academy performance. Although the TAV Selection System had both significant and high multiple correlation co-efficients, these co-efficients were not cross-validated, nor does the time between collection of predictor and criterion measures allow accuracy in determining if the validity co-efficients were concurrent or predictive. Nevertheless, this program pointed out the predictability of academic performance.¹²

f. A background information evaluation study by McAllister 1970 compared performance evaluation of New York City recruits during the first 18 months of their careers with initial judgment of the background investigations. Performance evaluations were based on seven criteria: time lost, time lost for injuries, completion of training standards, formal recognition of outstanding performance, absence of formal disciplinary charges, completion of probationary period and evaluation by supervisors. The study concluded that the background investigation was not highly valid. Further, those identified to be of higher character did not achieve higher performance than those with fair to poor character evaluations. Despite fair to poor character evaluations, recruits were certified eligible as a result of the total selection

¹²

Morman, Robert R.; Heinkey, Richard O.; Heywood, H. L.; Liddle, R., TAV Selection System, Police 1966 10(3) 54-58

process and their successful performance did not correlate on the basis of background investigation results alone.¹³

g. Selection tests employed by the US Army in determining eligibility for the Military Police Corps are the same tests as any soldier takes. These are Armed Forces Qualification Test (AFQT) and Army Classification Battery (ACB). Such tests are designed to measure aptitude and trainability rather than job performance predictability, emotional control and initiative. Military police are required to attain a minimum qualifying score of 100 (above average) in the ACB area considered most relevant for military police work (the skilled technical). This battery tests science knowledge (SK), arithmetic reasoning (AR) and mathematic knowledge (MK). Raw scores of these three areas are converted to an Army standard score. The AFQT percentile score equals the raw scores of three ACB tests (AR word knowledge and pattern analysis) summed and converted into a percentile score.¹⁴ These tests are not constructed on a job performance basis and screen no more than learned knowledge to date and prediction of general academic success. A more detailed explanation of these tests is at Appendix 1. A study of the Effectiveness of

¹³ McAllister, John A., Study of the Prediction and Measurement of Police Performance, Police 1970 14(4) 58-64

¹⁴ Psychological Testing in the US Army, US Army Research Institute for the Behavioral and Social Sciences, 1 Sep 73, pp 4-6

Selection and Classification Testing conducted by the US Army Research Institute for the Behavioral and Social Sciences (1973) revealed the following:¹⁵

- (1) Mental qualification is determined primarily on the basis of AFQT.
- (2) Additional evidence of trainability is obtained from aptitude scores on the AQB or ACB.
- (3) Tests are not constructed to determine job performance predictability of military police personnel as a separate selector.
- (4) Selection and classification tests used by the Army have a predictive accuracy that yields a validity co-efficient of between .5 and .6. (A co-efficient of 1.0 means that criterion performance is perfectly predictable from the test.)
- (5) Correlation between AFQT score and job training performance is about .5 on the average. The ACB has an average co-efficient of .65 with relevant training performance.

¹⁵ Effectiveness of Selection and Classification Testing, US Army Research for the Behavioral and Social Sciences, Sep 73

(6) Test correlations are more valid for the combat arms requirements than special skill areas.

(7) Means of Negroes on tests are well below those of Caucasians. This fact has suggested that lower prerequisites for job training or minimum qualifying scores be established (establish separate norms for each racial group and select from each group separately).

(8) Present tests when applied to performance referenced training predictors would be expected to have little or no validity in predicting level of performance.

Results of the above study focus on a problem in using the ACB as a screening device for US Army law enforcement personnel. If 1,000 potential military police were selected by a test that had no validity and the top 500 scores are selected, 250 of the selected men would be above the norm (performance average of the population) and 250 below. With a perfectly valid test, all of the top 500 selected would be above the norm. If a test with a validity co-efficient of .6 were used to select the top 500 men, 352 or 70 percent would perform above the norm.¹⁶ The remaining 30 percent may or may not be trainable to the desired level or may be carried through and assigned to protect a limited accession figure.

¹⁶ IBID

h. Departments of the Air Force and Navy were queried as to what psychological tests, if any, are employed in their selection process for law enforcement personnel. The Navy response, to include Marines, indicated use of general aptitude tests and cross-entry practices for periods of temporary duty. The Marines offer a more extended career field approach for enlisted personnel but no psychological test screening. The Air Force, which is 80 percent physical security duty and only 20 percent law enforcement mission related, uses general testing similar to the Army. Their requirement for a score of 30 on the battery tests equates to the Army's 100 on the ST score. In recent months the Air Force has increased interest in selection techniques for air policemen. A study similar in part to this one has recently been tasked to the Human Resources Laboratory, Air Force Systems Command, Lackland AFB, Texas. The task is so new that no directions or conclusions are available.

i. Thus far, it can be seen that as more sophisticated statistical methods have been employed and the present validation model moved from concurrent in present police officer prediction to recruit predictive attempts, the validity co-efficients decreased from the .60's to below .10. It is very important to note that while most results were not very encouraging, the higher multiple correlations obtained when data were analyzed separately by race are significant in both legal and practical terms.

3. Nonvalidity Studies.

a. This section, unlike the last which dealt with studies describing the relationship between predictors and criteria, reviews some of the studies which have focused on predictors without an attempt to assess validity. These studies are oriented to determining general characteristics of policemen as measured by tests or to assess relationships among a number of predictor measures. Since the validity question weighs so heavily in the use of psychological testing, the tests addressed in this section were given only superficial review. Those who desire more in-depth information will have to explore the original studies.

b. Nowicki (1966) administered the Minnesota Multiphasic Personality Inventory to police officers and industrial workers who were matched according to age, sex and educational level. Results reflected differences between the groups on a number of scales. Other differences were found between officers tested and the population used to develop the MMPI norms.¹⁷

c. Rhead, et al (1968) worked with the Chicago Police Department screening applicants who had both an extreme profile on the MMPI and marked disturbances in projection of body image on the Draw-A-Person Test. The study, over a two-year period (N=1,000), uncovered only difference in degree, not quality on the MMPI. There were differences found on the D-A-P

¹⁷ Nowicki, Stephen, Jr., A study of personality characteristics of successful policemen, Police 1966 10(3), 39-41

Test but they were not statistically significant.¹⁸ Although the results claimed to differentiate between chronic pathology and usable aggressive energy, no specific data were offered to support this conclusion.

d. In 1969, Gottesman used the MMPI on police candidates (N=203) who had successfully passed departmental selection procedures and male veterans (N=100) who received educational and vocational counseling from the VA. Results revealed that both groups were homogeneous within their own groups. Both policemen and veterans were significantly different from MMPI norms. The police exhibited fewer and less significant deviations than the veterans to the MMPI norms. On the basis of his study, Gottesman questioned the validity of using the MMPI norms as a basis for police profile comparison.¹⁹

e. A battery of tests administered to recruits in New York Police Academies included a Civil Service Examination, Lorge Thorndike Test, Police Background Information Test, Personality Inventory, Police Opinion Questionnaire, Police Knowledge Test and Discretionary Situations Test. Candidates were divided into two groups according to whether they passed or failed the Civil Service Examination. These two groups were then compared on performance on the other tests in the battery. A number of factors were

¹⁸ Rhead, Clifton; Abrams, Arnold et al, The Psychological Assessment of Police Candidates, American Journal of Psychiatry, 1968, 24(11), 1575-1580

¹⁹ Gottesman, Jay, Personality patterns of worker police applicants. Hoboken, NJ Laboratory of Psychological Studies, Sep 69

found to significantly differentiate between recruits who passed and those who failed the Civil Service Examination. Passed candidates were less religious, had more college education and related job experiences than failed candidates. Failed candidates were more strongly religious and had a greater number of siblings, unskilled fathers and months unemployed prior to application. Passed candidates were also more realistic than failed candidates in terms of the amount of danger they associated with certain situations and the maximum sentence they assigned to certain crimes. Failed candidates held a lower opinion of various services required of a policeman that are "not in the book" (i.e. assist stranded motorists, deliver babies, etc.).²⁰

f. Blum, et al (1960, 1961) conducted a study of selection procedures for deputy sheriffs in San Mateo, California. Data for this study consisted of correlations between oral board decisions and a number of test results including the Rorschach Ink Blot Test, Minnesota Multiphasic Personality Inventory, The F Test, Draw-A-Person Test, Test of Practical Judgment, Army General Classification Test, Strong Vocational Interest Blank and O'Rourke General Adaptability Test for Policemen. An essay question, life history summary, motivational essay and observation of behavior were also used. In the 1960 study, the Civil Service Examination was used as a preliminary screening device. Thus, all candidates in this study had passed the Civil

²⁰ Goldstein, Leo S., Perspectives on Law Enforcement Final Report, Princeton, NJ: Educational Testing Service, Feb 71

Service Examination. In addition, Oral Board members were not given access to supplemental test findings. Some of the major conclusions from this study were as follows: (1) Strongly unfavorable information existed on 42 percent of the men passed by the Oral Board; (2) Certain of the tests (e.g. Minnesota Multiphasic Personality Inventory) supposedly provided information that was not detected by other means. According to the authors, other sources of information (i.e. Army General Classification Test, O'Rourke General Adaptability Test for Policemen, Draw-A-Person Test, the observation accuracy and report writing test) did not seem useful enough to merit further study; (3) Of those men who passed the Civil Service Examination, those who were subsequently failed by the Oral Board had lower civil service scores than those who were subsequently passed by the Board; (4) All candidates found to have severe emotional disturbances were failed by the Board, but nearly all those with moderate disturbances were passed; (5) There was no relationship between Oral Board decisions and test findings of "character defect" (e.g. dishonesty, immorality, dangerous impulsivity, etc.) or use of judgment and intellect; and (6) The Minnesota Multiphasic Personality Inventory performed with some efficiency in detecting character defect. Evidence of such defect from MMPI profiles was substantiated by findings on other sources in seven out of eight cases. On the other hand, the MMPI failed to detect dishonesty in three out of nine men known to be dishonest.²¹ Blum's second study (1961) was basically the same as the first

²¹ Kent, D. A.; Eisenberg, Terry, The Selection and Promotion of Police Officers: The Police Chief, Feb 72

except for two major changes. First, the Civil Service Examination was employed but not as a preliminary screening device. Candidates continued in the program regardless of the Civil Service Examination scores. Secondly, the Oral Board members had access to supplemental findings prior to final actions on the candidates. The following six conclusions of the second study were deemed significant:²²

(1) Previous Civil Service Examination failures were certified eligible when the examination was not used as a prescreening device.

(2) Civil Service Examinations correlated .70 with OTIS IQ scores.

(3) Providing supplemental findings to the Oral Board further refined the detection of individuals with potentially undesirable qualities.

(4) The background investigation correlated closer with the Oral Board than other evaluation devices.

(5) Personality tests, motivational essay and number of jobs held were most consistently nonrelated to final pass or fail results.

(6) In 1960 the psychological tests agreed with psychiatrist's evaluation of 66 percent while later tests showed only a 33 percent agreement.

²² IBID

4. Present State of the ART and Conclusions.

a. From the standpoint developing psychological testing methods to be used in selection of law enforcement officers, the research done to date is rated as poor with one exception. Most studies reviewed have failed to address the major issue of differential predictive validity. Numerous studies dealing with identification of various test performance characteristics are of limited value in the search for effective and unbiased police selection. Validation has been addressed only in terms of correlation between one predictor measure and another. A predictive setting test of the evidence of concurrent validity has not been accomplished. Lastly cross-validation procedures and studies are so rare that usable techniques in this area are almost unavailable.

b. Studies in this review contained conclusions which were not supported by findings or were confusingly ambiguous. The absence of cross-validation, numerous predictor variables, and/or concurrent validity seriously questions the ability of most studies in this section to meet and support legal validity requirements of Title VII of the Civil Rights Act discussed in a later section of this study.

c. There is a bud of evidence supporting the potential prediction value of some written objective and situational tests in predicting police officer performance, but additional research is required to assess this potential.

d. To date, the only solid evidence for consistent predictive validity rests with relationships between some psychological tests of intelligence and aptitude and measures of police academy performance.

e. Review of paper and pencil psychological tests reveals that only current studies by The Industrial Relations Center of the University of Chicago (developed in detail in a later part of this study) bear any further study value in developing a psychological testing technique for use by the US Army in the selection of military police personnel.

f. None of the studies reviewed applied the factors of test construction based on job requirements and minority discrimination elimination.

g. For future research in this area a bibliography for this review is at Incl 2.

SELECTION FOR ENLISTMENT

PROGRAM	Standard Screening and Accessioning of Male Regular Army Enlistment Applicants
APPLICANT POOL	Male applicants <u>without prior military service</u> who must be screened for mental acceptability prior to enlistment (AR 601-210)
PLACE OF TESTING	Armed Forces Examining and Entrance Stations (AFEES)
TESTS ADMINISTERED	<p>Army Classification Battery (ACB-1973), Test Booklets 1 and 2, Form A and A-1; Form B for retest.</p> <p>For men enlisting under an enlistment option plan requiring International Morse Code Aptitude: Auditory Perception Test (AP) DA Form 6175, and ACB Test Booklets 1 and 2.</p>
SCORING	<p>ACB-1973 tests and scoring formulas: Manual, DA Pma 611-70; DA PT 4946</p> <p>a. Test Booklet 1, Form A or A-1, DA Form 6170 or 6174. Manual, DA Pamphlet 611-70-1. (All tests scored Rights only).</p> <ul style="list-style-type: none">(1) Mechanical Comprehension (MC)(2) Arithmetic Reasoning (AR)(3) Word Knowledge (WK)(4) Mathematics Knowledge (MK)(5) Pattern Analysis (PA) <p>b. Test Booklet 2, Form A or A-1, DA Form 6172 or 6176. Manual, DA Pamphlet 611-70-2. (All tests scored Rights only).</p> <ul style="list-style-type: none">(1) Automotive Information (AI)(2) Trade Information (TI)(3) Science Knowledge (SK)

clnd 1 b5c

- (4) Attention-to-Detail (AD)
- (5) Classification Inventory (CI)
- (6) General Information (GI)
- (7) Electronics Information (EI)
- c. Auditory Perception (AP), DA Form 6175
(formerly Radio Code Aptitude Test,
ARC-1) scored R - W/2

Aptitude Area Composites. ACB-1973 Test Booklet 1 and 2 test raw scores in each of ten aptitude area composites (see below) are summed. The raw score sum for each composite is converted to an Army Standard Score. The following raw scores are obtained from the Classification Inventory (CI): Maintenance (CM), Attentiveness (CA), Electronic (CE), Combat (CC).

- a. CO (Combat) = TI + AD + CC + AR + PA
- b. FA (Field Artillery) = CA + GI + EI + AR + MK
- c. EL (Electronics Repair) = TI + CE + EI + MC + AR
- d. OF (Operators & Food) = AI + CA + GI
- e. GM (General Maintenance) = AI + SK + MC + AR
- f. MM (Motor Maintenance) = AI + TI + CM + EI + MK
- g. CL (Clerical) = AD + CA + AR + WK
- h. ST (Skilled Technical) = SK + AR + MK
- i. GT (General Technical) $\frac{1}{2}$ = AR + WK
- j. SC (Surveillance/Communication) = MC + AR + WK + PA

AFQT Percentile Score. The raw scores of 3 ACB tests (AR, WK, PA) are summed. The raw score sum is converted to a percentile score. Manual, DA PT 4943

1/
GT (ACB-1973) is not used for selection and classification of male applicants.

QUALIFYING STANDARDS

Male Enlisted Applicants

a. high school graduates

one aptitude area score of 90 or higher
other than GT^{1/}

b. non-high school graduates

two aptitude area scores of 90 or higher
other than GT^{1/}

Accession processing: On the basis of aptitude area scores and interview information (civilian education, occupation, avocation and hobbies, and other background material), interviewer discusses with applicant and makes recommendation for direct MOS award, MOS for training, or enlistment option plan.

Enlistment Options^{2/}

- a. Army Service School Enlistment Option--3 or more aptitude area scores of 90; qualifying score in the prerequisite aptitude area established for the school course for which applying. If prerequisite qualifying score is 90 or higher, it may be considered as one of three 90's.
- b. Army Career Group Enlistment Option--Qualifying standard score as prescribed by DA Pamphlet 350-10 for aptitude area corresponding to ACGP for which applying; standard scores of 90 or higher in ^{3/}at least two additional aptitude areas.
- c. Army Airborne Enlistment Option--Aptitude area score of 80 in Aptitude Area CO.
- d. Army Special Forces Enlistment Option--see page 62.
- e. United States Army Security Agency (USASA) Enlistment Option--Aptitude area score of 100 in Aptitude Area ST.^{3/}

^{1/}GT (ACB-1973) is not used for selection and classification of male applicants.

^{2/}ACB-1973 Aptitude Area GT is not used for selection for training in MOS producing school courses for individuals tested under ACB-1973. GT will continue to be used for assignment to various programs and organizations (AR 614-200) except as changes are made to substitute a different aptitude area.

^{3/}For MOS or ACGP with an International Morse Code Aptitude requirement an AP Standard Score of 100 or higher is prerequisite.

- f. United States Army Special Intelligence Enlistment Option--Aptitude area score of 110 in Aptitude Area ST.
- g. Army Language School Option--See page 64.
- h. Army Ranger Enlistment Option--Aptitude area scores of 100 in Aptitude Area CO.
- i. Army Medical Skills Enlistment Option--Aptitude area score of 100 in Aptitude Area GM for Brace Specialist, Dental Laboratory Assistant, Optical Laboratory Apprentice.

Aptitude Area score of 100 in Aptitude Area ST for Dental Assistant, Clinical Specialist, Dental Hygienist, Physical Therapist, Occupational Therapy Assistant, X-Ray Assistant, Pharmacy Specialist, Veterinary Animal Specialist, EENT Specialist, Medical Laboratory Assistant, Senior Medical Laboratory Specialist, Diet Cook, Preventive Medicine Specialist.

- j. Army Engineer Skills Enlistment Option--Aptitude Area score of 90 in Aptitude Area GM for Carpenter, Mason, Electrician.

Aptitude Area score of 90 in Aptitude Area MM for Engineer Equipment Repairman.

Aptitude Area score of 90 in Aptitude Area GM for Crawler Tractor Operator, Crane Operator or Crane Shovel Pile Driver Operator, Grader Operator.

Aptitude Area score of 95 in Aptitude Area ST for Construction Draftsman.

Aptitude Area score of 100 in Aptitude Area ST for Construction Surveyor.

Aptitude Area score of 100 in Aptitude Area GM for Plumber, Heating & Cooling Specialist.

- k. Army Officer Candidate School Option--See page 36.
- l. Army Warrant Officer Flight Training Option--See page 72.

- m. Army Training and Travel Enlistment Option--AFQT score of 31; minimum applicable aptitude area score for appropriate MOS or ACG; qualifying Aptitude Area score for school or ACGP may be counted as one of the three aptitude area scores required when it is at least 90^{3/}.
- n. Army Strategic Communications Command Enlistment Option--Aptitude Area scores of 90 on three Aptitude Areas; if overseas area of choice desired, score of 100 in Aptitude Area CL or EL.
- o. Army 3d Infantry Enlistment Option--(The Old Guard) Aptitude Area Score of 90 in Aptitude Area CO.

^{3/}For MOS or ACGP with an International Morse Code aptitude requirement an AP Standard Score of 100 or higher is prerequisite.

BIBLIOGRAPHY
FOR
A REVIEW OF PSYCHOLOGICAL TESTING
IN POLICE SELECTION

- Abbatiello, A. A study of police candidate selection. Paper presented at the seventy-seventh annual convention of the American Psychological Association, Washington, DC, 1969.
- Baehr, McLany E., Furcon, John E., and Froemel, Ernest C., Psychological assessment of patrolman qualifications in relation to field performance. Washington, DC; US Government Printing Office, 1968
- Blum, Richard H. The polygraph examination in law enforcement personnel selection. *Police*, 1967, 12(2), 60-75.
- Blum, Richard H., Goggin, William L., and Whitmore, Earl. A study of deputy sheriff selection procedures: Preliminary report. Unpublished manuscript. May 1960.
- Blum, Richard H., Goggin, William L., Whitmore, Earl, and Pomeroy, Wesley A. A further study of deputy sheriff selection procedures. Unpublished manuscript. April 1961.
- Colarelli, Nick J. and Siegel, Saul M. A method of police personnel selection. *Journal of Criminal Law, Criminology and Police Science*, 1964, 55(2), 287-289.
- DuBois, Philip H. and Watson, Robert I. The selection of patrolmen. *Journal of Applied Psychology*, 1950, 34(2), 90-94.
- Gambol, Anne Marie, Slowikowski, Norbert S., and Doyle, Chet. Detective selection gets a new twist in Chicago. *Public Personnel Review*, 1965, 26(1), 40-43.
- Goldstein, Leo S. Perspectives on law enforcement: I. Characteristics of police applicants. Final Report. Princeton, New Jersey: Educational Testing Service, February 1971.
- Gottesman, Jay. Personality patterns of urban police applicants as measured by the MMPI. Hoboken, New Jersey: Laboratory of Psychological Studies, September 1969.
- Hill, Glen A. Potentials of Civil Police Psychological Tests in Selection of Military Police. Unpublished thesis. University of California, 1962.
- Hooke, James F. and Krauss, Herbert H. Personality characteristics of successful police sergeant candidates. *Journal of Criminal Law, Criminology and Police Science*, 1971, 62(1), 104-106.

Kaplan, H. Psychological Testing Progresses in the US Army. US Army Research Institute for the Behavioral and Social Crimes, 1 Sep 1973.

Kent, D.A., Eisenberg, Terry. The Selection and Promotion of Police Officers. The Police Chief, Feb 1972.

Levy, Ruth J. Predicting police failures. Journal of Criminal Law, Criminology and Police Science, 1967, 58(2), 265-276.

Levy, Ruth J. Investigation of a method for identification of the high-risk police applicant. Berkeley, California: Institute for Local Self-Government, 1970.

Maier, M.H. and Fuch, E.F. Effectiveness of Selection and Classification Testing US Army Research Institute for the Behavioral and Social Sciences, Sep 1973.

Matarazzo, Joseph D., Allen, Bernadene V., Saslow, George, and Wiens, Arthur N. Characteristics of successful policemen and firemen applicants. Journal of Applied Psychology, 1964, 48(5), 123-133.

McAllister, John A. A study of the prediction and measurement of police performance. Police, 1970, 14(4), 58-64.

McConnell, William A. Relationship of personal history to success as a police patrolman. Unpublished master's thesis, Colorado State University, January 1967.

Mills, Robert B., McDevitt, Robert J., and Tonkin, Sandra. Situational tests in metropolitan police recruit selection. Journal of Criminal Law, Criminology and Police Science, 1966, 57(1), 99-104.

Morman, Robert R., Hankey, Richard O., Heywood, Harold L., and Liddle, L. Rogers. Predicting state traffic officer cadet academic performance from theoretical TAV selection system scores. Police, 1966(a), 10(3), 54-58.

Morman, Robert R., Hankey, Richard O., Heywood, Harold L., Liddle, L. Rogers, and Goldwhite, Marie. Multiple prediction of municipal police officers' ratings and rankings using the theoretical TAV selection system and certain non-test data. Police, 1967, 11(3), 19-22.

Morman, Robert R., Hankey, Richard O., Kennedy, Phillis K., and Jones, Ethel M. Academy achievement of state traffic officer cadets related to TAV selection system plus other variables. Police, 1966(b), 10(6), 30-34.

Narrol, Harvey G. and Levin, Eugene E. Formal assessment procedures in police selection. Psychological Reports, 1963, 12, 691-694.

Newman, James M., Rogin, William E., Hunter, William K., and Vora, Shanti. Investigation of a method for identification of the high-risk police applicant: An analysis and interpretation based on available data. In Can high-risk police applicants be identified? Berkeley, California: Institute for Local Self-Government, July 1971, 59-89.

Nowicki, Stephen, Jr. A study of the personality characteristics of successful policemen. Police, 1966, 10(3), 39-41.

O'Connor, George W. Survey of selection methods. Washington, DC: Field Service Division, International Association of Chiefs of Police, 1962.

Reinke, Roger W. and Eisenberg, Terry. National police registry and testing service conference -- Colorado Springs, December 10-12, 1970: Summary. Washington, DC: Professional Standards Division, International Association of Chiefs of Police, 1970.

Reinke, Roger W. and Eisenberg, Terry. Police personnel management: A selected bibliography of recent literature. Gaithersburg, Maryland: Professional Standards Division, International Association of Chiefs of Police, November 1971.

Rhead, Clifton, Abrams, Arnold, Trosman, Harry, and Margolis, Philip. The psychological assessment of police candidates. American Journal of Psychiatry, 1968, 24(11), 1575-1580.

Spencer, Gilmore and Nichols, Robert. A study of Chicago police recruits: Validation of selection procedures. Police Chief, 1971, 38(6), 50-55.

Zaice, Joseph E. Measured interests, personality, and intelligence of professional policemen. Unpublished master's thesis. Washington State University, 1962.

D

STUDIES
BY
THE INDUSTRIAL RELATIONS CENTER
UNIVERSITY OF CHICAGO

1. Introduction.

a. The finest validity studies in the area of psychological testing in selection of law enforcement personnel were initiated in 1966 by the Industrial Relations Center at the University of Chicago in cooperation with the Chicago Police Department and funded by a grant from Office of Law Enforcement Assistance, US Department of Justice. These three agencies represented a professional police department willing to participate in such a study, an institution staffed by scientists with the knowledge and equipment to perform such a study, and financial support. Studies to date represent several projects. The first project from 1966 to 1968 attempted to identify selection tests, selection standards and types of patrolman performance. The second project from 1969 to 1971, a follow-up of the first, judged job performance measures, investigated performance predictions, examined racial group differences and analyzed performance level over a three-year period. The third project from 1971 to 1973 established as a basic objective the development of a low cost, systematic and effective procedure for selecting police officers and placing them in assignments best utilizing their mental abilities, skills and behavioral attributes.

b. This section of the review will be addressed in three units. First will be a general discussion of the three projects. The second will address the type of pencil and paper tests employed in these projects. The third will list conclusions drawn from the entire review.

2. General Discussion.

a. The first project conducted during the period 1966 to 1968 involved an extensive battery of commercially available written tests administered to two groups of police officers (N=490) who had experience of one year or more. Eight different performance measures were employed. Two of these were subjective (a paired comparison of supervisory ratings and a departmental performance rating) and six were objective (tenure, awards, complaints, disciplinary actions, arrests and times absent). In this project, very high and significant multiple correlations were reported between test battery performance and all eight performance or criterion measures of job performance. Multiple correlations were reported in the .50's and .60's with the two types of supervisory ratings as performance criteria. When the multiple correlations were corrected for statistical bias, co-efficients were reduced to .40's and .50's.¹ Cross-validation resulted in a further lowering of multiple correlations to the .30's and .40's. During this project, only

¹Baehr, Melany F., The Appraisal of Job Performance. Occasional paper Nr 27.41, Chicago, IL. Industrial Relations Center, Univ of Chicago, 1968

concurrent validity is described.² Results of the first project, limited by time and follow-up techniques, were not very encouraging in terms of validity. One of the inherent problems was that concurrent validation was employed rather than predictive. As discussed in the review of psychological testing in Police Selection portion of this study, evidence supporting concurrent validation does not support predictive potential.

b. The second project initiated in 1968 was a follow-up on the first project. This project included four principle research goals:³

(1) Determination of the stability of the main supervisory and objective measures of police performance utilized in the 1966-68 study. The continuation study permitted an evaluation of the adequacy of performance measures used in the 1966-68 study based on subsequent patrolman performance. It also allowed data collection concerning peer ratings of patrolman performance.

(2) Investigation of predictive validity of the psychological test batteries used in the 1966-68 study. Comparison of predicted performance levels of patrolman, made on the basis of their 1966 test results with subsequent job performance, could yield information about the success of the test battery in predicting future performance levels of police officers.

² Baehr, Melany F., Psychological Assessment of Patrolman Qualifications in Relation to Field Performance, LEAA, Project #46, US Dept of Justice, 5 Nov 68

³ FURCON, et al and assessments of patrolman field performance, Industrial Relations Center, University of Chicago, Chicago, IL, 1972

(3) Examination of the nature and stability of racial group differences found in response to test batteries from the 1966-68 study. Verification of differences in response to test batteries, by black and white officers, compared with job performance differences was carried out to provide guidelines for the most effective selection of applicants of both races.

(4) Conduct of a longitudinal study of performance levels of the eight patrolman subgroups identified in the 1966-68 study. This enabled the gathering of additional evidence on the stability and significance of these previously identified groups.

The follow-up study addresses performance reviews of 1966 candidates in the 67, 68, 69 period. Introduction of patrolman peer rating is discussed and indications of a favorable acceptance by the department are presented. Additional information on racial group differences in response to the test battery and the issue of equitable and effective selection procedures is presented. The relationship between predicted and actual performance is established. The follow-up longitudinal study reported significant predictive multiple correlations ranging from .13 to .27. These correlations were obtained by assessing relationships between 1966 police officer battery test performance and actual job performance as much as three years later, measured by seven of the eight performance criteria utilized in the first study. Finally, in this follow-up project, using newly appointed patrolmen, the highest predictive multiple correlation between test battery performance

and department supervisory ratings covering a subsequent period of 2 1/2 years was only .07.

c. To summarize the first two projects, it is shown that as more sophisticated statistical procedures were employed and the validation model moved from concurrent through police officer predictive to recruit predictive, the validity co-efficients shrank from the .60's to less than .10. Despite apparent grim results there were some important findings. For instance, although a co-efficient of .27 was found between 1966 test battery results and actual performance in 1969 as measured by the paired comparison supervisory rating, co-efficients of .34 and .42 were found when police officers were separated into black and white groups respectively. This finding adds further support to the requirement for performing differential validity studies. Cross-validation procedures also indicated that the best prediction of performance was obtained when the weights, based on a specific racial group, were applied to members of the same group, particularly to the black group. Conversely, predictors of performance were the poorest when weights from one racial group were applied to members of the other group.

d. The third project was a cooperative police selection test validation project undertaken by the Illinois Association of Chiefs of Police, The Illinois Association of Boards of Fire and Police Commissioners, and the Industrial Relations Center of the University of Chicago. The goal of the project was validation of specific psychological tests for use in selecting

police officers in local communities throughout the state and establishment of general standards for use in the police selection process.⁴ Basic steps involved in the project are as follows:⁵

(1) Analysis of Basic Job Requirements, - to determine personal qualifications of the local community police officer necessary for occupational success.

(2) Selection of Psychological Tests - to identify tests thought to measure key qualifications necessary to success as a police officer.

(3) Establishment of Job Performance Criteria - to develop supervisory evaluations and identify objective performance data available in municipal department to serve as the standard against which to gauge effectiveness of the tests.

(4) Selection of Patrolman Sample - representative in state.

(5) Test Administration - to currently employed police officers.

⁴ FURCON, John and Fromemel, Earnest C., The Relationship of Selected Psychological Tests to Measures of Police Officer Job Performance in State of IL. The Industrial Relations Center, Univ of Chicago, Chicago, IL, 15 Oct 73

⁵ IBID

(6) Statistical Analysis and Validation - identify psychological tests which show a relationship with on-the-job performance of police officers.

(7) Development of an Occupational Standard - based on (representative) police test performance.

An application of police chiefs, better performing officers and poorer performing officers Skills and Attribute Inventory produced the following essential skills or attribute list of personal qualifications required to be in effective police officers:⁶

(1) Interpersonal Effectiveness - the capacity to act effectively in extremely divergent interpersonal settings.

(2) Questioning and Listening Skills - the ability to quickly and effectively obtain information and sort out relevant from irrelevant.

(3) Capacity for Analysis, Judgment and Problem Solving - the ability to make practical, common sense decisions; to develop plan of action quickly; to help people solve problems; to anticipate actions of others.

(4) Emotional Stability and Control - the ability to control his emotions and impulses in the face of extremely provocative conditions, to withstand

⁶ IBID, page 5

verbal and physical abuse, to act effectively and calmly when his own life may be in jeopardy.

(5) Orientation of Service to Other Persons - to perform routine tasks, such as escorts or intersection control, with a positive attitude; to take positive action in crime or accident prevention.

(6) Perceptual Skills - the ability to "size up" or grasp situations quickly, to perceive departures from routine events or differences from an established pattern.

(7) Decision Making Capacity - the ability to make a prompt decision and see that it is carried out.

(8) Leadership and Direction of Others - the ability to act as a leader, to assume control of ambiguous, upset or problem situations involving people.

(9) Persistence - not giving up when a solution or remedy is not readily apparent.

(10) Flexibility and Versatility - to deal effectively with the task at hand in spite of rapid changes in pace of activity, in the situation itself or in the types of persons involved.

(11) Communications Abilities - to make effective oral reports, to write clear and concise descriptions of incidents in preparation of formal reports, to properly complete required administrative forms.

(12) Alertness - to maintain perceptual vigilance and readiness to respond during long periods of routine or even monotonous activity.

(13) Psychomotor Skills - to effectively perform psychomotor activities required of a police officer.

(14) Objectivity and Personal Integrity - to render service in a professional manner, put aside personal feelings or attitudes, remain objective in all dealings, and exhibit a high level of personal integrity and ethical conduct.

Tests were then selected to assess the above qualifications (tests selected are described in the next unit). Tests were administered to a varied number of police officers (N=563) representing large, medium and small police departments from Chicago metro, north and south Illinois. Results were analyzed in relationship between test scores and patrolman job performance. Results showed multiple correlation of .37 for state-wide application with higher correlations in the Chicago metro area .52, northern region .55, and southern region .83. Test battery results, with limited data on black

officers, showed promise in removing many artificial or cultural barriers surfaced in the review of other selection programs.

3. Tests Selection for the Industrial Relations Center Projects.

a. Based on results of the occupational analysis, approximately 40 different standardized tests were reviewed for possible use in the projects. All of the tests were thought to have at least some possibility of assessing qualifications in areas identified as significant in the occupational analysis. The final selection of tests was made with the following objectives in mind:⁷

(1) To develop a test battery with strong potential for predicting successful police officer performance.

(2) To develop a test battery which could be readily adopted for use by Illinois Police and Fire Commissions in screening of police applicants.

A number of other technical and practical considerations were taken into account.

(1) Emphasis on group tests - only tests which could be administered by trained clerical personnel on a group basis were included. This eliminates the need for a highly skilled professional staff in the initial screening step.

(2) Emphasis on paper and pencil tests - this eliminates the need for special apparatus or equipment.

(3) Emphasis on objective scoring - only tests which could be directly scored with a prepared key were included, eliminating the need for subjective interpretation.

A summary description of tests selected for use in the project is presented below. A more detailed description of each test is presented at Incl 1.

b. Summary Description of Tests.

(1) Tests of Motivation

(a) Personal Background Inventory

-untimed, approximately 25 minutes to complete

-a biographical questionnaire which presents both factual and opinion questions about the respondent's background and gauges the strength of motivational drives based on personal achievements.

(b) Work Interest Index

-untimed, approximately 15 minutes to complete

-an occupational interests inventory which measures strength of interest in various occupational areas as well as breadth of interests and level of aspiration.

(2) Tests of Intellectual Skill

(a) Nonverbal Reasoning

-untimed, approximately 25 minutes to complete

-measures reasoning capacity through the solution of pictorial problems.

(b) Understanding Communication

-timed, 15 minutes

-measures ability to read and comprehend written material.

(c) Listening Comprehension

-timed, approximately 30 minutes

-assesses the respondent's oral comprehension skills in areas of immediate recall of information, following oral directions, recognizing transitions in oral communication, and recognizing word meaning.

(d) Perceptual Speed

-timed, 5 minutes

-measures the ability to rapidly identify similarities and differences, to detect discrepancies in visual material.

(e) Closure Speed

-timed, 3 minutes

-measures the ability to rapidly organize apparent unrelated visual material into meaningful wholes, to accurately define and organize incomplete visual data.

(f) Closure Flexibility

-timed, 10 minutes

-measures the ability to identify visual figures despite distraction.

(3) Tests of Aptitude

(a) Cassel Test of Social Insight

-untimed, approximately 30 minutes

-assesses an individual's capacity to "size-up" problem situations involving the behavior of people in group settings.

(b) Chapin Test of Social Insight

-untimed, approximately 25 minutes

-measures a person's ability to analyze problem situations involving a single individual's behavior.

(4) Tests of Behavior

(a) Arrow-Dot Test (IRC Modification)

-timed, 6 minutes

-an "objective" and performance test of personality which assesses the degree and type of impulse control shown by the individual.

(b) Press Test

-timed, three parts of 90 seconds each

-measures reaction speed and performance under simple conditions and stress conditions induced by the test.

(c) Temperment Comparator

-untimed, approximately 20 minutes

-a self-descriptive test of personality dealing with factors of introversion-extroversion, degree of emotional control, self-reliance, energy level and sociability.

4. Conclusions.

a. That the methodology used for these projects represents important insights to future police selection.

b. That evidence was reported that some tests were less likely to discriminate unfairly against minority groups.

c. That as each project refined the battery of test used and supervisory rating techniques, the predictability potential increased.

d. That the tests were developed and/or selected based on an analysis of the job requirements.

e. That the technique involved in these projects can be applied to other communities in other states. (The Industrial Relations Center is currently undertaking a police selection study in the City of Detroit.)

DETAILED DESCRIPTION OF TESTS
USED IN THE PROJECT

TESTS OF MOTIVATION

PERSONAL BACKGROUND INVENTORY
[PBI]

AUTHORS: Melany E. Baehr, Ph.D., and Frances M. Burns,
Industrial Relations Center, The University of
Chicago.

MEASURES: This inventory assists in predicting future job success on
the basis of past performance and experience. Past
performance and experience are evaluated by an analysis
of the scores obtained on 10 performance factors. These
performance factors are derived from 94 actual items of
information concerning family, education, and work
history; they were identified through a series of factorial
studies. Their names are given below:

Group Leadership and School Achievement

Drive

Mobility

Financial Responsibility

Family Responsibility

Job and Personal Stability

Educational-Vocational Consistency

Parental Family Adjustment

Professional-Successful Parents

General Health

SAMPLE ITEMS:

-
- | | |
|------------|--|
| 1. Do you: | a. <input type="checkbox"/> Own a house
b. <input type="checkbox"/> Rent a house or apartment
c. <input type="checkbox"/> Rent a room
d. <input type="checkbox"/> Live with relatives or friends
e. <input type="checkbox"/> Other |
|------------|--|
-
- | | |
|---|--|
| 2. While in high school, where did your
grades rank you in your class? | a. <input type="checkbox"/> Upper 10%
b. <input type="checkbox"/> Upper 25%
c. <input type="checkbox"/> Upper 50%
d. <input type="checkbox"/> Lower 50%
e. <input type="checkbox"/> Don't know |
|---|--|
-

TIME LIMITS: Untimed. Approximately 20 to 25 minutes.

TESTS OF MOTIVATION

WORK INTEREST INDEX (Non-Verbal) [WII]

AUTHORS:

Melany E. Baehr, Ph.D., Richard Renck, Ph.D.,
and Robert K. Burns, Ph.D., Industrial Relations
Center, The University of Chicago.

MEASURES:

This Index is pictorial and completely non-verbal. It consists of 96 pictures, each of which shows a man engaged in a specific occupation. For each picture the subject checks whether he would like "L" or dislike "D" the work shown. The principal purpose of the Index is to assist in determining an individual's pattern of occupational interests through an analysis of the scores he obtains on 12 work interest factors and two work attitude scales (Aspiration Level and Flexibility of Interests). The 12 factors were defined through a series of factorial studies. They are named and interpreted as follows:

Professional & Technical. Professional and technical skills used for therapeutic purposes--physical therapist, dentist.

Social & Verbal. Social and verbal skills used in teaching or counseling and resulting in the esteem of others--school principal, office manager.

Authority & Prestige. Authority and prestige or independent judgment used in unpredictable or non-routine situations--civil lawyer, newspaper reporter.

Artistic & Interpretative. Artistic and interpretative talents used in the performing arts--actor, composer.

Artistic & Stylized. Artistic ability used to make stylized products--taxidermist, glassblower.

Artistic & Creative. Artistic or creative talent used in the communication of ideas--sculptor, commercial designer.

Technical & Scientific. Technical and scientific skills used in the physical or biological sciences--physicist, pharmacist.

Clerical & Routine. Routine clerical skills used in concrete procedures--general office worker, inventory clerk.

MEASURES:
cont.

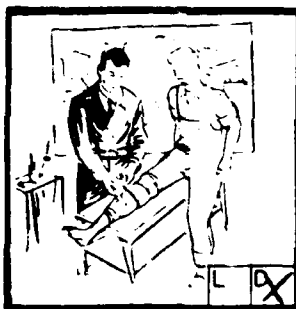
Business Contact & Structured. Business contact in structured social situations--hotel desk clerk, grocery checker.

Personal Service & Persuasive. Personal service for the advantage or enjoyment of others--ship's purser, welfare director.

Mechanical & Productive. Mechanical skills for producing or repairing a product--automobile mechanic, cabinet maker.

Control of Massive Equipment. Direction of massive equipment in largely masculine settings--locomotive engineer, cable driller.

SAMPLE ITEMS:



TIME LIMITS:

Untimed. Approximately 10 to 15 minutes.

TESTS OF INTELLECTUAL SKILL

NON-VERBAL REASONING [NVR]

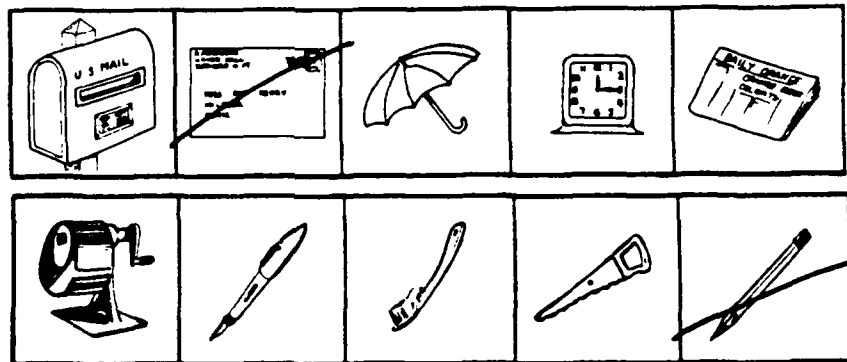
AUTHOR:

Raymond J. Corsini, Ph.D., Industrial Relations Center,
The University of Chicago.

MEASURES:

This test measures reasoning ability through the medium of pictorial problems. Each of the 44 test items consists of a row of five pictures. The subject is asked to find 'the picture that goes best with the first picture in the row.' He then draws a line through the picture he selects. The test yields a single score.

SAMPLE ITEMS:



TIME LIMITS:

Untimed. Approximately 25 minutes.

TESTS OF INTELLECTUAL SKILLUNDERSTANDING COMMUNICATION
[UC]AUTHOR:

T. G. Thurstone, Ph. D., The School of Education, The University of North Carolina.

MEASURES:

This test measures comprehension of verbal material in the form of short sentences and phrases. Vocabulary and speed of reading are of minor importance in achieving a good verbal comprehension score on it. Test items are incomplete sentences with four alternative completions. The subject selects the word or phrase which correctly completes the sentence.

SAMPLE ITEMS:

When waves hit against rocks, small pieces of the rock are broken off. When these broken pieces rub against other rocks, they break up again, making

A. rains. (B) sand. C. waves. D. floods.

If the pilot of an airliner meets too much ice, he can ascend or descend to different flight levels until he finds a warmer

A. engine. B. transport. (C) temperature. D. day.

TIME LIMITS:

15 minutes.

TESTS OF INTELLECTUAL SKILL

BROWN-CARSEN LISTENING COMPREHENSION TEST (Form AM) [B-C]

AUTHORS:

James I. Brown, University of Minnesota, and
G. Robert Carlsen, University of Texas.

MEASURES:

This test measures the ability to comprehend spoken language. By listening comprehension is meant the aural assimilation of spoken symbols in a face-to-face speaker-audience situation, with both oral and visual cues present. This form is comprised of 76 items which measure the following important listening skills:

Immediate Recall--which measures the ability to keep a sequence of details in mind until a question is asked which requires thinking back over the sequence.

Following Directions--which measures the ability to follow oral directions.

Recognizing Transitions--which measures awareness of the function of transitional words and phrases within sentence contexts.

Recognizing Word Meanings--which measures the ability to recognize meanings of words from context.

SAMPLE ITEMS:

Part A. Immediate Recall:

"In the series of numbers 8-7-6-3-2, the fourth number is ____."

Part B. Following Directions:

"In the upper left corner of your answer sheet is a group of numbers and letters to which you will need to refer." (Point to proper place.) "The even numbers and the vowels have been circled. Now look at the sample. Referring to the row of numbers above the answer spaces for this part, add the smallest number to the largest number. The answer is ____?"

1 ② 3 ④ 5 ⑥ 7 ⑧

① B C D ⑤ F G H

(from upper left
corner of answer
sheet)

SAMPLE ITEMS:

cont.

Part C. Recognizing Transitions:

"An introductory sentence sets the stage for discussion; a transitional sentence marks a change of thought or a new point; and a concluding sentence adds a note of finality to what has been said. 'It is indeed a privilege to address you on this occasion.' If you think the sentence is introductory, fill in the answer space under I; if you think it is transitional, fill in the answer space under T; if you think it is a concluding sentence, fill in the answer space under C; and if it is none of these, fill in the answer space under N."

Part D. Recognizing Word Meanings:

"In the sentence, 'The scouts pitched their tents,' which meaning best defines the word pitched?"

- a. set up b. threw
- c. furnished d. arranged
- e. fixed at a particular level

TIME LIMITS:

- Timed. Part A. Approximately 6 minutes.
 Part B. Approximately 6 minutes.
 Part C. Approximately 3 minutes.
 Part D. Approximately 3 minutes.

TESTS OF INTELLECTUAL SKILL

PERCEPTUAL SPEED (Identical Forms) [PS]



















AUTHORS:

L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D.,
The Psychometric Laboratory, The University of North
Carolina.

MEASURES:

This test measures the ability to identify likenesses and differences in visual configurations (i. e., diagrams, drawings, and figures). This ability has been identified as a primary one in visual thinking. It requires a minimum of thinking in the usual sense, and has very little to do with sharpness of eyesight. The subject is asked to indicate which of five similar figures is identical with a figure on the left. The test yields a single score.

SAMPLE ITEMS:

	1	2	3	4	5	
						1
						3
						

TIME LIMITS:

5 minutes.

TESTS OF INTELLECTUAL SKILL

CLOSURE SPEED (Gestalt Completion) [CS]

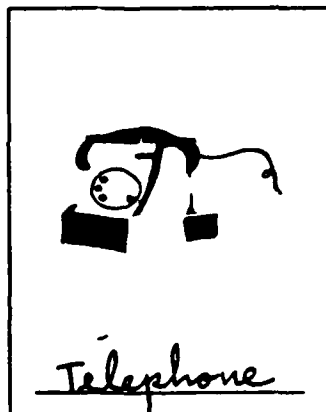
AUTHORS:

L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D.,
The Psychometric Laboratory, The University of North
Carolina.

MEASURES:

Closure speed is the ability to see an apparently disorganized or unrelated group of parts or ideas as a meaningful whole or single unit. In other words, it is the ability to unify a complex situation. In this test, the subject is asked to identify objects which are only partially or sketchily pictured.

SAMPLE ITEMS:



TIME LIMITS:

3 minutes.

TESTS OF INTELLECTUAL SKILL

CLOSURE FLEXIBILITY (Concealed Figures) [CF]

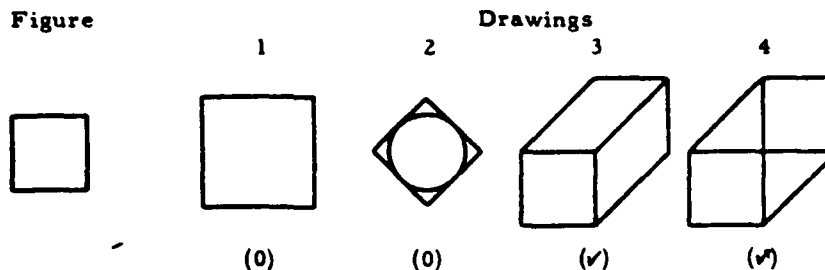
AUTHORS:

L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D.,
The Psychometric Laboratory, The University of North
Carolina.

MEASURES:

This is an objective test of the ability to keep a figure in mind in the face of distraction, that is, to see a given figure (diagram, drawing, or configuration of some kind) which is "hidden" or embedded in a larger, more complex, drawing, diagram, or configuration. The subject is given a figure followed by four drawings, and asked to indicate in which of the drawings the figure is "concealed." (The figure must retain its size and general spatial orientation.)

SAMPLE ITEMS:



TIME LIMITS:

10 minutes.

TESTS OF APTITUDE**TEST OF SOCIAL INSIGHT (Adult Edition)**
[TSI]**AUTHOR:**

Russell N. Cassel, Ed.D.

MEASURES:

This test appraises the characteristic mode of reaction an individual exhibits in dealing with interpersonal or social problems. It consists of 60 items, each of which describes an interpersonal problem and offers five alternative actions to take in handling it. The five alternatives cover the following five modes of resolving such problems:

Withdrawal--avoiding or escaping the problem by leaving the situation.

Passivity--remaining in the situation but taking no action to solve the problem.

Cooperation--initiating an active endeavor to solve the problem.

Competition--drawing the attention of others, attempting to deal with the problem by excelling or outdoing others.

Aggression--directing little effort to the problem but expressing hostility or strong feelings in response to it.

The frequency with which an individual selects each mode establishes his characteristic pattern in dealing with social and interpersonal problems.

SAMPLE ITEM:

After waiting in a busy garage for 20 minutes for the serviceman to take your repair order, he returns and starts taking the order of another patron who just arrived. What do you do?

- a. Report the error to the service manager.
- b. Insist that he take your order immediately.
- c. Walk out to your car and drive away.
- d. Say nothing and wait until he notices you.
- e. Explain to the serviceman that you arrived before the other patron and had been waiting for him to return.

TIME LIMITS:

Untimed. Approximately 30 minutes.

TESTS OF APTITUDETHE CHAPIN SOCIAL INSIGHT TEST
[CHAPIN]

AUTHOR: F. Stuart Chapin, Ph.D., University of Minnesota.

MEASURES: This test measures an individual's ability to appraise others, to sense what they feel and think, and to predict what they may say and do.

SAMPLE ITEM: Mr. Carpenter, when told that an acquaintance had purchased a pleasure boat, criticized him strongly for spending so much money for a boat when he probably could not afford one. Several days after this incident, Mr. Carpenter himself bought an expensive new pleasure boat. About the same time, he negotiated a second-mortgage on his house. Why did Mr. Carpenter criticize his acquaintance for an act he afterward performed himself?

- ☐ a. His acquaintance was probably an unsafe boat operator.
- ☐ b. Because Mr. Carpenter probably received an inheritance following the death of a relative.
- ☐ c. Criticism of his acquaintance served to release his feelings of uneasiness about something he planned to do himself.
- ☐ d. Second-mortgages are not uncommon for people of Mr. Carpenter's education and income.

TIME LIMITS: Untimed. Approximately 25-30 minutes.

TESTS OF BEHAVIOR

ARROW-DOT TEST-IRC Modification [A-D]

AUTHORS: Lawrence A. Dombrose, Ph.D., and Morton S. Slobin, Ph.D.

MEASURES:

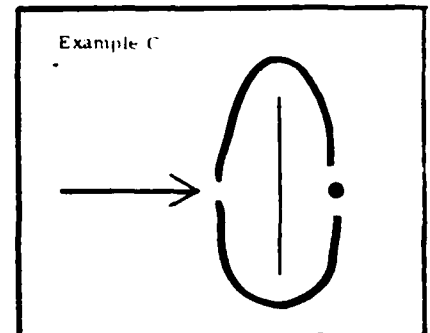
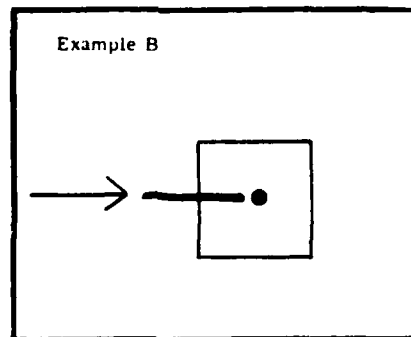
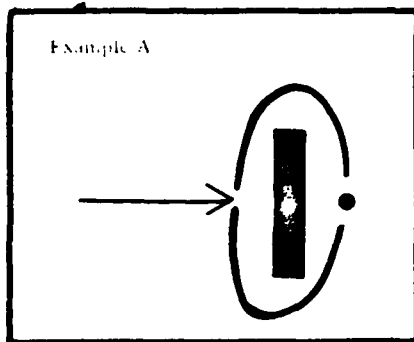
This projective test is an adaptation of one part of a four-part battery called The IES Test. Its 30 items each set a relatively simple perceptual-motor task--to draw a line from the point of an arrow to a dot. Arrow and dot are separated by a number of barriers in the form of solid or dotted lines and black bars. The instructions explain how various combinations of these are to be handled. The test yields three scores:

The I Score--a measure of the tendency to uncontrolled impulsive behavior through which needs are satisfied immediately and directly without any regard for realistic restrictions or moral issues.

The E Score--essentially a measure of the ability to function realistically, of the degree of ego strength and control.

The S Score--a measure of superego strength, especially in circumstances where the demands of the superego are not experienced consciously but are viewed as external restrictions.

SAMPLE ITEMS:



TIME LIMITS: 6 minutes.

TESTS OF BEHAVIOR

THE PRESS TEST [PRESS]

AUTHORS:

Melany E. Baehr, Ph. D., and Raymond J. Corsini, Ph. D.,
Industrial Relations Center, The University of Chicago.

MEASURES:

This is an objective test of personality, developed to measure an individual's ability to work under stress. It yields three main scores, one from each of its three parts:

- Part I. Speed of reaction to verbal stimuli
Part II. Speed of reaction to color stimuli
Part III. Speed of reaction to color stimuli in
 a stress situation caused by the interference of distracting verbal stimuli.

Two further scores which can be derived are the figures for Part I minus Part II and for Part II minus Part III. Directions for each part, with sample items, are given below.

SAMPLE ITEMS:

Part I

Put the first letter of each word in the corresponding circle.

1	2	3	4		1	2	3	4
Red	Blue	Green	Yellow		(R)	(B)	(G)	(Y)
Blue	Yellow	Red	Green		(B)	(Y)	(R)	(G)

Part II

Put the first letter of the colors of the dots in the corresponding circles.*

1	2	3	4		1	2	3	4
BLUE	YELLOW	RED	GREEN		(B)	(Y)	(R)	(G)
GREEN	BLUE	YELLOW	RED		(G)	(B)	(Y)	(R)

- *the color designations in sample are provided only for purposes of explanation; in the actual test booklet only the colored dots themselves are presented.

SAMPLE ITEMS:
cont.

Part III

Write the first letter of the color of the ink in which the word is printed in the corresponding circles.

1	2	3	4	1	2	3	4
Red	Blue	Green	Yellow	(R)	(B)	(G)	(Y)
Blue	Yellow	Red	Green	(B)	(Y)	(R)	(G)

TIME LIMITS:

Part I 90 seconds.
Part II 90 seconds.
Part III 90 seconds.

TESTS OF BEHAVIOR

TEMPERAMENT COMPARATOR [TC]

AUTHORS: Melany E. Baehr, Ph.D., and R. W. Pranis,
Industrial Relations Center, The University of
Chicago.

MEASURES: The Temperament Comparator measures the relatively
permanent temperament traits of the individual. It
yields measures of 18 temperament traits and of five
temperament factors, all within the normal range of
behavior. The traits are:

Calm	Persevering
Cautious	Prompt Starter
Decisive	Quick Worker
Demonstrative	Seeks Company
Emotionally Stable	Self-Confident
Energetic	Serious
Enthusiastic	Socially at Ease
Even-Tempered	Steady Worker
Lively	Talkative

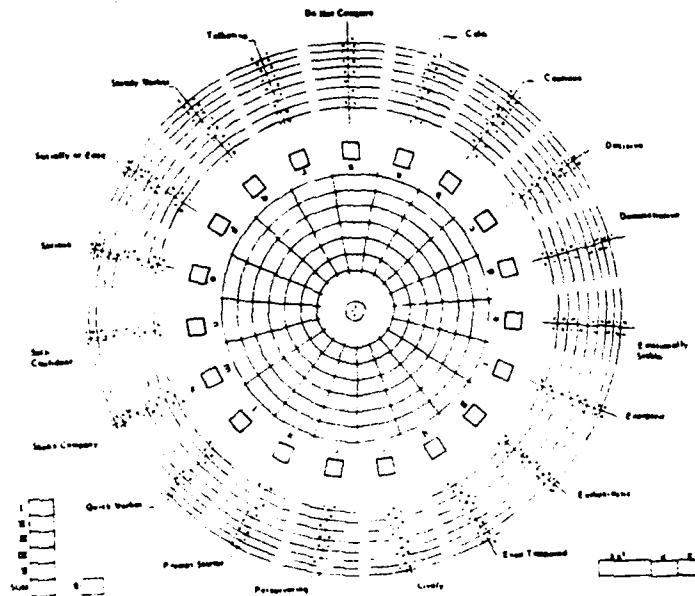
The five temperament factors based on these traits are:

Controlled
Stable
Self-Reliant
Excitable
Sociable

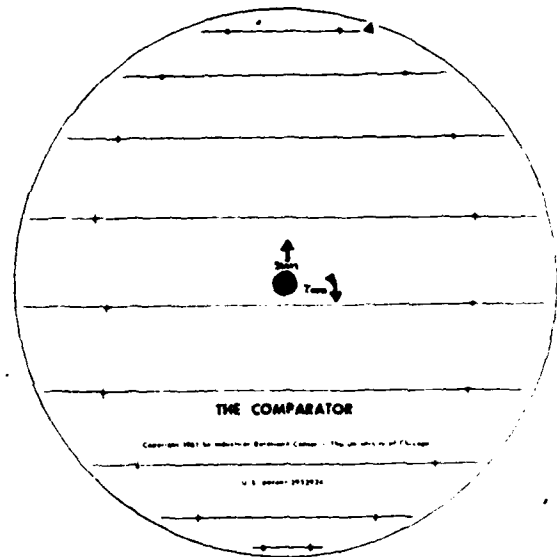
The items in the Comparator consist of all possible pair-
ings of the 18 temperament traits. The subject rotates a
disc to successive positions within a circle which has the
traits printed on its circumference. Bars on the disc link
pairs of traits. The subject examines each pair of traits
and places a mark in a hole on the disc near the trait which
he has selected as more descriptive of his behavior. The
Comparator also yields a Coefficient of Consistency which
indicates the extent to which the subject was consistent in
his selections or judgments.

SAMPLES:

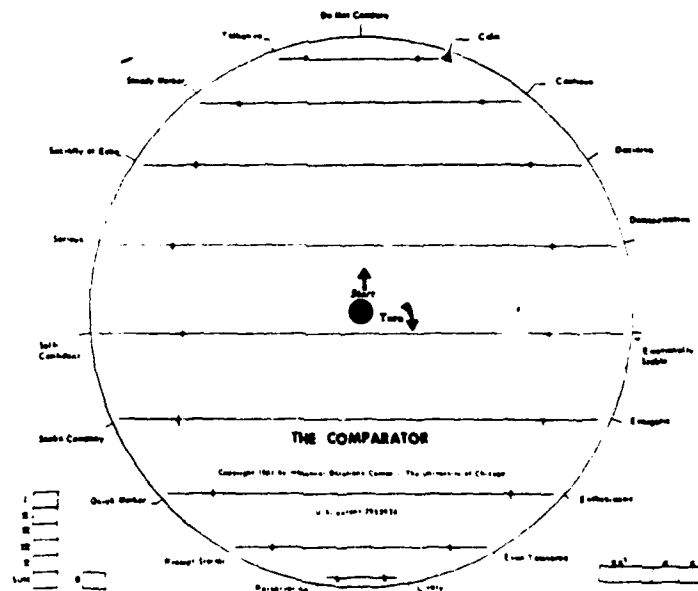
Circle with Temperament Traits



Cardboard Disc



Circle and Disc Assembled

**TIME LIMITS:**

Untimed. Approximately 15 to 20 minutes.

REFERENCES

- Baehr, Melany E. The appraisal of job performance (with special reference to the use of the Comparator). (Occasional Paper No. 27) Chicago, IL: Industrial Relations Center, The University of Chicago, 1963.
- Baehr, Melany E. The appraisal of job performance (results of five years' use of the paired comparison performance appraisal for the validation of selection test batteries). (Occasional Paper No. 27-R1) Chicago, IL: Industrial Relations Center, The University of Chicago, 1968.
- Baehr, Melany E., Furcon, John E., and Froemel, Ernest C., Psychological assessment of patrolman qualifications in relation to field performance. Washington, DC: Superintendent of Documents, US Government Printing Office, 1969.
- Furcon, John, Froemel, Ernest C., Franczak, Ronald G., and Baehr, Melany E. A longitudinal study of psychological test predictors and assessments of patrolman field performance. Chicago, IL: Industrial Relations Center, The University of Chicago, 1971.
- Furcon, John. Some questions and answers about police officer selection testing. (Occasional Paper No. 35) Chicago, IL: Industrial Relations Center, The University of Chicago, 1972.
- Furcon, John, Froemel, Ernest C. The relationship of selected psychological tests to measures of police officer job performance in the state of Illinois, Industrial Relations Center, University of Chicago, Chicago, IL, 15 Oct 73.
- Heckman, Robert W., Groner, Dennis M., Dunnette, Marvin D., and Johnson, Paul D. Development of psychiatric standards for police selection. Minneapolis, MN: Personnel Decisions, Inc., 1972.
- Tierney, John B. and Biggins, Peter A. The comparator (instrument, 1961; test administration manual, revised 1967). Chicago, IL: Industrial Relations Center, The University of Chicago.

A SUMMARY OF THE ASSESSMENT CENTER APPROACH

1. Background.

a. The term "assessment center" was introduced by Professor Henry Murray of Harvard University in his studies on personality research in the 1930's. The origin of multiple assessment procedures on a wide scale is credited to German military psychologists who selected officers by putting candidates through intensive assessment procedures. The British also used like procedures for screening their officers.

b. During WW II, the United States Office of Strategic Services (OSS) ran a series of assessment centers for the selection of their agents. The OSS, in a series of three day assessments, evaluated the leadership ability of 5,391 recruits. In addition to interview material, biographical data, and casual observations, the OSS staff gathered ratings on ten major traits in an average of six different situations per trait. The OSS approach illustrates the first serious consideration by psychometricians of the need to take into account particular situations in which individuals operate. The situational tests developed by OSS were designed to test each prospective's qualifications in a number of different job-related situations.

c. The prototype for present day assessment centers was developed from American Telephone and Telegraph "Management Development Program," which

was started in 1956.¹ Douglas Bray, and others of AT&T, used assessment center programs to learn about the backgrounds, skills, and abilities of new supervisors so they could gain insight into management development process and identify variables related to success.

d. According to Guyton, the assessment center concept of identifying executive or management potential is based on the following hypothesis:²

(1) There are differences in the personal characteristics of more successful and less successful executives.

(2) These differences can be identified.

(3) The chances are better for a candidate's success if his personal characteristics are more like those of the more successful executive than those of the less successful.

(4) These characteristics can be measured early in his career.

1. Bray, D. W. & Grant, D. L., "The Assessment Center in the Measurement of Potential for Business Management Psychological Monographs, 1966, 80 (17 whole no 625)

2. Guyton, T., "The Identification of Executive Potential." Personnel Journal, 1969, 866-872.

e. In the prediction of executive potential three steps are employed. First is the selection of criteria of executive success; second is the identification of characteristics associated with successful executives; third is the developing of instruments to validly measure predictors of success.

2. Assessment Center Description.

a. A typical assessment center is located centrally in the needs of the firm agency, etc. It is staffed by one or more professional psychologists and by temporary assistant evaluators drawn from the management personnel of the organization itself. The number of evaluators per group of persons to be evaluated may vary with the complexity of the task. Normally, the average, based on the various centers reviewed, in this study is three to four evaluators per every six persons evaluated, and either six or twelve persons to be evaluated are assigned in a given session. The length of a session will vary from one to five days, depending on the policy or task of the individual assessment center. The persons to be evaluated ("assesseees") report to the center and are put through a series of exercises designed to measure their management aptitude. The four categories of exercises are:

(1) Group simulations in which six candidates work together in an enterprise or activity.

(2) Individual tests.

(3) Individual simulations.

(4) An in-depth interview usually structured.

b. During the evaluations simulation assessees are placed in hypothetical management situations and required to perform management functions.

c. From the review of assessment center programs conducted by over 25 of the 100 companies and governmental agencies using this technique, group exercises appear to fall into two distinct categories. The first is the assigned role where the individuals are assigned a particular role usually competitive in nature. The second is the nonassigned role where individuals are free to assume any role within the limits of the structured exercise. During an assigned role exercise group participants might be assuming the role of a supervisor charged with the responsibility of selecting someone to fill a managerial vacancy from the ranks. Each individual is given data on one nominee and within a time limit must sell that nominee to the group within the best of his ability. The group must then select one nominee from all presented.

d. The most universally applied individual exercise is called the In-Basket Exercise. Twenty-four of the 25 centers studied used this exercise. An example of an in-basket exercise is shown at Incl 1. The in-basket contents, left by the assessor's predecessor in a management position, contains memos, letters and reports of varying importance. The assessor is limited in time and asked to document each action taken, including notes, letters and phone calls. At the completion of the exercise each candidate is interviewed by an evaluator to ascertain the reasoning behind his decisions, methods of arranging his work, and other information providing insight into his capabilities.

e. Approximately one-half of the centers surveyed used the background interview in their center operation. Interviews, conducted by an evaluator, last between one-half hour to two hours. A questionnaire completed by the assessee is used as a general guide during the interview. In addition, the interviewer asks standardized questions which relate to variables which are measured as part of the assessment process.

f. Observation and/or rating techniques vary in design but fall into two general categories, quantitative and non-quantitative. Quantitative data lends itself better to statistical interpretation but non-quantitative descriptive data appears more easily understood and a better tool for management personnel. Approximately half of the center programs reviewed employ pencil and paper psychological tests as a small but

integral part of the overall evaluation. The types of tests used are covered in the chapter on a review of psychological testing. In all cases the data is usually prepared by the evaluator in the form of a final written report in each assessee's performance at the center. One approach gaining greater preference is that of having the evaluator prepare narrative reports and a professional psychological staff translating these reports into usable qualified data for validation and other statistical uses. Reports or observations are normally prepared immediately after the assessment. An overall rating of high, medium or low is normally included. The final reports are sent to management personnel (supervisors, personnel departments, etc.) for use.

g. Approximately half of the centers' programs reviewed employ immediate feedback techniques. Of those centers employing feedback, all use at least oral feedback to the assessee. Several used an additional written followup in addition to the oral feedback. The feedback is given the assessees only by center personnel in some cases, while in other cases, it is given through line organization personnel.

3. General Discussion.

a. Validity tests on assessment centers to date have been less than pure. Despite the fact that most large participating organizations have operated their centers for sufficient time to collect validity data, the

center output has been fed back into the organization causing varying degrees of criterion contamination. An impressive validity study by AT&T (N=5,943) followed advancement in management. The criterion was advancement above first-level management. Assessment results were used as a basis for promotion to first-level management only, and the criteria (higher than first level) was felt to be relatively free of bias. Individuals assessed "above acceptable," were twice as likely to be promoted two more times than individuals assessed "as acceptable," and ten times more likely to be promoted than those assessed "not acceptable."³ In 11 studies on its assessment center, IBM found a positive relationship between center findings and various criteria of success.⁴ A comprehensive review of validity studies by three leading authorities (Byham, Cohen, and Moses) is at Incl 2.

b. Military studies in this area have been limited and the most complete work to date focuses on the assessment center pilot program at Fort Benning, GA. This program, staffed by 24 individuals, operates three day pilot testing in the selection for the OCS program, advance course and basic NCO's program and resource requirement determination.

3. Greenwood, J. M. and McNamera, W. J., "Leadership Styles of Structure and Consideration and Managerial Success." Personnel Psychology 1968, 21-85-98.

4. IBID

Like civilian centers, the Benning center employs psychological testing (personality tests and differential officer battery). Situational tests to demonstrate potential leadership behavior to solve tasks, and in-depth interviews (by combat arms officers). The combined assessment rating considers all of the above evaluators. To date, validity studies at Benning have been sketchy as feedback in times of unusual permanent change of station activity has been difficult. Since the Benning center is combat arms oriented, an in-depth study was dropped in favor of the above review.

c. The cost of center operations has varied in terms of money and manpower available. Costs per individuals assessed have been found to range from less than \$5.00 to approximately \$500.⁵ The differences developed from such factors as number of participants, staff numbers and salaries, length of assessment, exercise costs and facility operational costs. The Bell System 60 assessment centers, which operate with nine permanent staff members per center, produce highly reliable results at a cost of \$400 per candidate assessed. The Internal Revenue Service reported a cost of about \$275 per assessee in their pilot program.⁶ On the basis of

5. Byham, W. C., "The Assessment Center on an Aid in Management Development," Training and Development Journal, Dec 7.

6. Di Costanzo F. and Andretta, T. The supervisory assessment center in the Internal Revenue Service Training and Development Journal, 1970, 12-15.

the pilot study, the IRS decided to continue assessment on a slightly expanded basis. A realistic figure derived from this review would appear to be between \$250 and \$400 per person assessed. In terms of professional staff requirements, the size depends on the individual assessment procedures and the use of professional psychologists as assessors. Centers report as few as two and as many as ten employees. Each center has, as a minimum, one staff member who is professionally trained in the administration and interpretation of psychological tests. This person is also responsible for validation and various other research activities. A requirement exists for an administrator to assume responsibility for the operational and business contact duties. Secretarial and clerical employees are required for typing, administering and scoring tests, filing and distributing reports, tabulation, statistical analysis and related duties. The staffing guidance appears to be that of sufficient in size to expediently handle the workload.

4. Summary. Industrial and governmental organizations have recognized the need to identify and utilize management potential at an early stage. New and more complete information gathering devices are being developed to help management personnel make promotion decisions. The assessment center is gaining in acceptability as a promising technique for that end. By objective evaluation in a number of simulation problems, administration and interpretation of paper and pencil tests, and in-depth personal interviews, the assessment center personnel attempt an overall evaluation of

each individual. The narrative or quantitative report is used as a supplementary aid in determining promotion ability of individuals assessed. The assessment center provides the opportunity for an individual to demonstrate leadership or managerial talent early in their career. Only one of the centers studies addressed selection screening at entry level for law enforcement. The Mills study, not discussed thus far, appeared to be oriented toward possible use in the selection of military law enforcement personnel. This study is discussed in detail in the next section of this report. The use of techniques in this section is considered impractical for the Army law enforcement program without further costly modification and study.

5. For future research requirements, a complete bibliography is included as Incl 3.

THE IN-BASKET EXERCISE

In this exercise, the candidate is asked to assume the role of a manager several levels above his position.

The candidate is confronted with 20 to 30 items of varying importance (left in his predecessor's in-basket), and is given one to three hours to take appropriate action. Some typical items might be:

CITY ELECTRIC COMPANY

TO Fred:

March 23 1972
DATE

Sally Smith received a summons to serve on jury duty for three months, beginning May 1. She has indications that the trial load is heavy and that this will pretty well occupy her time for the jury term.

Now is the time to think about her replacement. Sally Peters shows the promise of moving in Sally Smith's job since she is our senior order typist. If we hired someone right into the Service Rep job, we will undoubtedly make Sally Peters unhappy. Shall I look into a new order typist now so that we can release Sally Peters for S. R. training?

Jan Wilson

CITY ELECTRIC COMPANY
30 NORTH STREET
JUNCTION CITY, USA

R. C. Cola, Manager
Silver City

Dear Sir:

The latest report of overtime hours for the five offices in your district shows

IND 1 TAB C

your office to be highest. It appears that running a "tighter operation" is called for.

Please let me know what steps you plan to reduce overtime hours. Could I have a reply by April 1.

If some action is not taken immediately, I will be forced to recommend a thorough investigation of your organization by our time and study people

Yours very truly,

C. A. Boss
District Manager

The assessee is also given a list of people occupying various positions throughout the company and a calendar for scheduling.

Time for the exercise is one half to three hours.

Summary of Validity Research

ABSTRACT

A review of all validity studies concerning the assessment center since 1956 revealed that the assessments were robust and were consistently related to several criteria. Predictive accuracy was highest for job potential, followed by progress, and then job performance. The assessment center was more efficient than traditional methods of assessment, such as paper-and-pencil tests. Various design problems regarding the validity studies were discussed and suggestions for future research and practice were given.

THE VALIDITY OF ASSESSMENT CENTERS:

A LITERATURE REVIEW

BARRY M. COHEN¹

University of West Florida

JOEL L. MOSES

American Telephone and
Telegraph Company

WILLIAM C. BYHAM

University of Pittsburgh

Since 1956 when the American Telephone & Telegraph Company introduced the process to United States industry on a widespread basis, assessment centers have markedly increased in usage and application. At present it is estimated that more than 60 companies now have assessment centers, over 100 companies are developing them, and hundreds of companies and governmental representatives have attended assessment center conferences in the last two years (Byham, in press). Interest in the assessment center method is widespread. For example, Bray and Moses (1972) suggest that the increased use of the assessment center method was one of the major new developments that was seen in their review of personnel research.

The assessment center is a comprehensive, standardized program in which participants are evaluated for selection, training, or career planning purposes. Multiple observational techniques are used and each participant is evaluated along a number of previously determined management dimensions. A team of assessors observes and evaluates each participant on the dimensions and

makes an overall judgment of each participant's potential for advancement, development or placement.

For purposes of this article, an assessment center program is defined by the following characteristics:

1. Multiple techniques, such as situational exercises, interviews, objective and projective tests, peer ratings, and other performance measures are used; at least some of the techniques require group interaction. The techniques are selected or designed to bring out performance relative to dimensions previously identified as important to job success.
2. From four to six trained assessors staff the program. Usually they are unfamiliar with one another and the assessees, do not directly supervise the assessees, and are two to three supervisory levels above them. Assessment training for the staff varies from one to three days to one month. A psychologist may serve on the staff or be the director of the program.
3. From six to twelve participants are assessed at the same time, and there is a low ratio of assessees to assessors (1-1 to 4-1) which allows systematic and close observation and multiple evaluations of each assessee by several assessors.
4. The assessment staff integrates and interprets its observations of each assessee judgmentally and may

use psychological tests or other psychometric measures to supplement its judgment. The strengths and weaknesses of an individual on each dimension are evaluated, followed by a global rating, which is an overall prediction based on assessment performance.

5. The assessment center is conducted off the job but not necessarily off company premises, usually in a 2 to 5 day period.

Assessment center programs also have unique features which vary with the organizations utilizing them. Programs may vary in particular techniques used, in their administration (e.g., scheduling exercises), in identifying and choosing assessees, in assessor training and in the role of the psychologist in the program (Byham, 1970).

This article is concerned with evaluating the overall assessment rating. This is a global rating in which the assessment panel judgmentally combines its observations of the assessee's performance throughout the assessment center and arrives at an overall prediction of subsequent job performance, progress or potential. Despite the many unique features of assessment center programs, the global rating results from a process common in all such programs. The process is three-fold:

1. determining the dimensions to be assessed by conceptualizing worker requirements for the positions studied

2. judging the assessee's standing on those dimensions from assessment center performance
3. combining these judgments into a global, overall prediction.

This judgmental process differs from clinical assessment where the clinician relies mainly on a global model of personality. The assessment global rating follows the identification of common performance dimensions.

The purpose of the article is to review and critically evaluate the research literature, including predictive and concurrent studies of assessment programs in business and government. The period of this review extends from 1956 to July, 1972, beginning with the Management Progress Study's use of an assessment center (Bray 1964b), the most well known and first industrial application of this method.

HISTORICAL PERSPECTIVE

Three separate periods can be seen which trace the development of assessment center applications. These are the Initial, the Research, and the Implementation periods concerning the use of the assessment center method.

The Initial Assessment Period introduced the use of multiple assessment methods and provided the foundation for the technology of assessment. The Research Assessment Period, beginning in the mid 1950's, adapted the methodology of assessment in industry, and provided much of the basic research needed to support its applied use. The Implementation Period, beginning in the late 1960's,

adapted the assessment center method on a widespread basis to a variety of settings, organizations and purposes, and resulted in a proliferation of research concerning assessment.

Assessment activities prior to 1956 constitute the Initial Assessment Period. Assessment techniques were used for a variety of purposes including validation of techniques, personality research, and selection (Taft, 1959), but generally were not well received for lack of empirical validity. Prior to World War II no validity studies were reported, despite sporadic assessment activity first conducted by German military psychologists in 1911 to select military officers, and later by Henry Murray (1938) in personality research.

Assessment center programs first caught the attention of psychologists during World War II, for the selection of high level military and support personnel. The British War Office Selection Board (WOSB) (Harriss, 1959; Morris, 1949), the U.S. Office of Strategic Services (OSS) (OSS Assessment Staff; 1948), and the British Civil Service Selection Board (CSSB) (Vernon, 1950), independently conducted such programs for military, secret service, and civil service officers, respectively. The CSSB program is still operational today, mostly in the same form as it was in 1945 (Moore, 1969). Generally, mixed validities were reported for these programs, but no study reported entirely invalid findings. For example, OSS validity coefficients ranged from .08 to .53 (OSS Assessment Staff, 1948), WOSB from .13 to .41 (Harriss, 1959; Morris, 1949), and CSSB from .26 to .49 (Vernon, 1950).

These programs tended to rely heavily on content validity in lieu of criterion-related validity because selection was urgent. Criteria were often inexact, unreliable, unattainable, or contaminated, which could explain the mixed validity findings.

After World War II, a small number of psychologists began programs to identify the career potential of professional trainees, graduate students and officer candidates using an assessment format. Again, mixed validity findings were reported. Low validity findings include the Kelly and Fiske (1951) study of clinical psychologists in VA hospitals in which objective tests and a credentials file were as predictive of clinical competence as the assessment staff; the Holt and Luborsky (1958) Menninger School of Psychiatry study which obtained validity coefficients of around .25; and the Institute of Personality Assessment and Research study (IPAR) of air force captains (MacKinnon, 1958) which obtained a median validity coefficient of .13. High validity findings include the Chicago Assessment Program (Stern, Stein & Bloom, 1956); the IPAR study of advanced graduate students (Barron, 1954; Gough, 1953) (assessment staff global rating correlated .41 with potential ratings); and the Officer Candidate School Program (Holmen, 1956) (validity coefficients of around .55).

Two articles reviewing the validity studies of the Initial Assessment Period, by Taft (1959) and Chronbach (1970), include analyses that may explain the mixed validity findings of the period. Programs of this period with high validities tended to differ from others which had less success in several ways. The

more successful programs had assessors who were familiar with the criterion task and the assessee's work experience. Assessment techniques included simulations or work samples, rather than relying solely on psychological tests, and assessors interpreted performance on the basis of task demands and not personality theory.

In the Research Assessment Period, the first widespread application of the assessment center method took place in industry, pioneered by A.T.&T. A longitudinal research program, the Management Progress Study, was initiated by psychologists at A.T.&T. in 1956. This study, which is still in progress today, was designed to follow the career development of a sample of young managers. In order to evaluate their initial potential as well as to establish benchmarks for longitudinal changes, an assessment center modeled after the OSS program was developed. The participants were assessed and re-assessed eight years later (Bray 1964b; Bray & Grant, 1966; Grant, Katkovsky, & Bray, 1967; Grant & Bray, 1969). Since the assessment results were used for research rather than for organizational purposes, they provided a rich source of data relating the contribution of various assessment techniques to management progress, as well as the stability and reliability of the method itself.

Within the Bell System, assessment centers became operational in 1958, and were designed for use in aiding in the selection of entry level managers. These first applied operational centers received considerable attention, and studies reporting on their effectiveness began appearing in the literature.

For example, Campbell and Bray (1967) compared the performance of individuals promoted following attendance at an assessment center with individuals promoted without assessment. Similarly, Bray and Campbell (1968) described a predictive validity study used to develop an assessment center for salesmen. Moses (1971) reported a recent follow-up on over 6000 participants in the Bell System program.

Unquestionably, these A.T.&T. studies have aroused much attention and support for the assessment center method. A large number of organizations, seeing the success of the Bell centers, began similar assessment center programs. The earliest of these adaptations began in 1963 at Standard Oil (Ohio). Since then validity studies of this program (Carleton, 1970; Thomson, 1970; Finley, 1970; Finkle & Jones, 1970), as well as those of other early centers such as IBM (Wollowick & McNamara, 1969; Hinrichs, 1969; Dodd, 1971; Kraut & Scott, 1972); General Electric (General Electric, 1969; Meyer, 1970); and Sears Roebuck (Bentz, 1967, 1971) have been reported.

These programs, as well as many others, represent the current assessment Implementation Period. The widespread application of assessment programs to many industries and governmental agencies as noted by Byham (1970), Bray and Moses (1972), and others indicate considerable acceptance of this method. Assessment programs designed to select managerial talent have been developed and used for all levels of management. In addition, numerous assessment programs have been designed for specific needs such as identification of training and development needs; placement; early identification of

talent; selection of supervisory, sales, and technical specialists; and identification of opportunities for such target groups as minorities and women. These applications, implemented in a wide variety of settings, characterize the current use of assessment center methods.

Advocates of the assessment center program have argued strongly that it is a highly valid method. Jaffee (1966, 1967) has suggested that the use of simulation exercises minimizes prediction error because the performance observed is very similar to task demands of the managerial role, and experienced assessors familiar with the role essentially need not make extended predictions in rendering a global rating. Dunnette (1971) argued that simulation is a highly valid means of assessing performance because the performance manifested is similar to the worker-organization interaction. Moses (1969) stated that the use of multiple assessment techniques is the most positive feature of the assessment center process because it results in high bandwidth or information variety from which a more accurate and dynamic prediction can result. Byham (1971) argued that the validity of assessment centers is due to reasons such as: the exercises often are job samples of the kinds of work expected if the individual is promoted; the setting is unusually conducive to behavioral observation (away from the work setting); the assessors give full attention to observation; and behavior is used to predict behavior. These writers conclude that the assessment center method is a more valid method for management selection than other traditional methods of assessment, such as supervisory appraisals or paper-and-

pencil tests.

Common Research Problems

In reviewing assessment validity studies, a number of common problems in research methodology were found. Among these, the possibility of criterion contamination has been the most common and serious problem. To varying degrees, line management has had the benefit of the assessment center reports for personnel decisions in all but two assessment center programs studied (Bray & Grant, 1966; Bray & Campbell, 1968). These programs may be called operational programs to distinguish them from research programs with uncontaminated criteria.

Operational programs have been subject to additional confounding effects, depending upon the research design employed. Experimental designs used to validate the process include the following:

1. Follow-up of assessed group. A group of subjects is chosen by various procedures including self, supervisor, or other nominations. Depending upon the nomination procedure, the assessed group might not be representative of the employee population from which it was chosen. In the minds of higher management, a positive or negative halo might be associated with the assessed group and result in additional criterion bias. This problem can be partially avoided by using various nomination procedures and choosing assessees from several geographical or organizational areas. In the smaller organization, however, these steps might not be feasible. Results

can be improved the longer the experimental group is followed up in a longitudinal manner.

2. Comparison of assessed and non-assessed groups. While more desirable than the follow-up method, this approach may add confounding to the extent that groups were not adequately matched on variables, which may affect assessment center or criterion performance. The initial participants selected for assessment in a new program may differ significantly from participants seen once the process has been accepted by the organization. As such, when subjects assessed early in the program's development are compared with a non-assessed group, validity findings may be distorted to the extent that the assessed group is not representative. On the other hand, waiting for a representative sample of assessed subjects may make comparisons with an older selection method more difficult.
3. Comparison of high performing and low performing assessment groups. This design can be used with both the follow-up and the assessed non-assessed comparison designs. Typically, progress in management is used as a dependent variable. Obtaining a sample of low rated assessees who are promoted despite their assessment performance is quite difficult. This approach, however, can be used in a longitudinal setting.

VALIDITY STUDIES

A summary of 18 research studies is presented in Table 1. These

studies, conducted from 1964 to July, 1972, represent research in a variety of industrial settings. Data on the type of study, the location, assessors, subjects and criterion measures used are presented.

- - - - -

Insert Table I about here

- - - - -

The most comprehensive and well-known study of an assessment center program to date is A.T.&T's Management Progress Study (Bray & Grant, 1966). This study compared the assessment center performance of 123 college and 144 non-college men and their subsequent progress in management. All were assessed by a special research assessment center staff shortly after reaching the first (entry) level of management. Each participant was assessed during the years 1956 to 1960. Neither the participant nor management was aware of the assessment results, and considerable effort was taken to prevent any possible contamination of the study information.

Comparisons made between the assessment center predictions and subsequent progress in management were highly significant. For example, the assessment staff made a global prediction, based on all of the assessment information, whether or not each subject would, in fact, reach middle management within ten years from the time of assessment. Management level obtained 5 to 8 years later, showed a highly significant relationship to the assessment predictions. For example, 42% of those seen as material for middle management had attained this level, while only 7% of the low rated assesseees pro-

gressed to this level. Similarly, in the high rated assessment group, only 4% still remained at the entry management level 5 to 8 years later, while 42% of the low rated group had not been promoted.

In addition to an analysis of the assessment center ratings, comparisons were made between the assessment techniques, their contribution to the overall assessment rating, and their relationship to progress.

Comparison between assessment performance and salary progression also showed a significant relationship. For example, the assessment overall rating correlated .48 (Mdn r for total assessed groups) with this criterion, and no other single technique (assessment trait ratings, situational exercises, or paper-and-pencil tests) correlated as highly. Since the global rating was based upon data from these techniques, this finding indicated that the assessment staff could successfully integrate this data judgmentally. A more important finding was that when tests were partialled out, the global rating continued to explain reliable variance of 10% to 18% in most samples. Thus, the assessment center method had justified its cost compared with the paper-and-pencil test variety of managerial selection.

Since the Management Progress Study, over 75,000 men and women have been assessed in Bell System operational assessment center programs to date, this being an excellent index of the receptivity of the assessment center method in that organization. Four follow-up studies of these programs have been completed. These studies help to determine the extent to which the assessment center, once operational

and affected by a multitude of administrative factors (i.e., staffing, location, assessor training), is effective. The question is, does validity generalize from research to operational programs?

The objective of the majority of A.T.&T. assessment programs was managerial selection, specifically of craftsmen qualified for first level management with potential for higher level management. An early study conducted in the Michigan Bell Company compared job performance and potential for advancement of an assessed group promoted to management with a group promoted before the assessment center program, but unfortunately the groups were not matched (Bray, 1964a). Results revealed that nearly twice as many high performance and potential men were found at first level management in assessed as compared to non-assessed groups. Another study conducted in the New England Bell Company (Mather, 1964) compared acceptable and not acceptable assessed groups consisting of craftsmen promoted to first level management and first level managers promoted to second level. Of the acceptable group, 76% were rated as showing "high quality job performance," while only 40% of the not acceptable group received this rating. Another follow-up study (Campbell & Bray, 1967), using four other Bell companies, compared three groups (N=223) of first level candidates assessed "acceptable," "questionable," and "not acceptable" and subsequently promoted to management, and two non-assessed groups (N=223) promoted before and after the assessment center program. As in the other follow-up studies, this study was also limited to recently promoted managers in the assessed group but, unlike the other studies, by denoting a

less than acceptable assessed group of those promoted, the false negative rate could be determined. Moreover, by studying two non-assessed groups, possible halo bias due to being promoted before versus after the assessment center could be determined.

While significant differences in job performance between those rated as "acceptable" by the assessment center and those rated as "not acceptable" were obtained, the largest differences obtained were between ratings of management potential. While half of the "acceptable" subjects were seen as having high potential for further advancement, less than one quarter of the not assessed group had further advancement potential. As a group, even those assessed and rated "not acceptable" and subsequently promoted had more potential than either group of non-assessed promotees. The assessment staff global ratings, thus, was most useful in predicting job potential.

A final, recent follow-up study (Moses, 1971) compared the progress of 5943 men who were assessed for subsequent career progress. A very stringent criterion measure (those attaining two or more promotions since assessment) was used. A correlation of .44 between the overall assessment rating and progress was obtained, again demonstrating considerable relationship between the overall rating and subsequent success.

Other companies have found that the assessment process works quite effectively in a variety of organizational settings. For example, researchers at Standard Oil (Ohio), using the basic assessment format developed at A.T.&T., found positive relationships between assessment performance and management progress criteria. From 1963

to 1966, 122 entry level management candidates were assessed and followed up for a $2\frac{1}{2}$ to 5 year period (Carleton, 1970). A later sample of 109 such candidates was assessed by the same program and followed up for a somewhat shorter period (Finley, 1970). A multitrait-multimethod validation study of the SOHIO program for both samples involved various measures. Criteria included 13 supervisory ratings averaged from two independent raters and a composite measure of managerial progress (salary growth and promotions, adjusted for initial status). Predictors included 13 ratings from each of three methods including projective test, interview, and assessment staff. The assessment staff method was not independent of the other methods, however. Staff ratings were based upon assessment reports including the other methods, paper-and-pencil tests, and sociometric data.

As in the Management Progress Study (Bray & Grant, 1966), the SOHIO validity study included a managerial progress criterion and, additionally, included 12 job performance ratings and a potential rating." Thus, this study's findings would reveal the relative usefulness of the assessment center program in predicting each of these criteria. The findings confirmed the earlier A.T.&T. findings: that is, the program's assessment staff ratings were moderately valid for the prediction of managerial performance (early sample, $\text{Mdn. } r = .25$; later sample, $\text{Mdn. } r = .32$) and highly valid for the prediction of potential (early sample, $r = .65$; later sample, $r = .63$) and progress (early sample, $r = .51$). These validity coefficients consequently exceeded those for performance ratings based upon the

interview (early sample, $r = .05$ to $.33$) and projective test (early sample, $r = .11$ to $.25$; later sample, $r = .01$ to $.34$).

In a related study Thomson (1970) used a multitrait-multi-method analysis to compare the performance of 71 managers evaluated by managers and psychologists at a SOHIO assessment center and subsequently evaluated by supervisors from 6 months to 2 years after attending the center. Thirteen ratings were made by assessors and supervisors. In general, the judgments made at the center showed a high degree of convergent validity between assessors (psychologists and managers) with a Mdn. r of $.85$ between the raters.

On the other hand, the supervisor's ratings were much poorer than the predictor ratings, which raises a question of the appropriate criterion measure for assessment, a theme to be returned to later. The supervisors were more lenient, used a more restricted range of ratings, and had lower reliabilities than ratings made by the assessors. This finding makes a strong argument against simply giving supervisors a list of predetermined assessment dimensions and asking them to rate their subordinates. Given the appropriate techniques and standardized procedures found at an assessment center, however, these same individuals can be expected to make meaningful judgments.

This was a major conclusion of an early IBM study reported by Greenwood and McNamara (1967). They compared the ratings made by psychologists and managers who participated in an assessment center. No significant differences in ratings were found between these groups, and the rater reliability in both groups was quite high.

Other IBM studies have demonstrated strong relationships between assessment center performance and management success. A study by Wollowich and McNamara (1969) reviewed the performance of a sample of 94 lower and middle managers 3 years after assessment. From job analyses of positions held, a criterion measure of progress (increase in management responsibility) was constructed, and this measure probably was somewhat more unbiased than others previously used, such as management level. Two predictor methods were used: the assessment staff global rating and a statistical weighting of performance data (tests, simulation exercises, and trait ratings), using linear regression. Using either method, predictions were valid, but the statistical weighting was considerably more efficient ($r = .62$) than the global rating ($r = .37$). The judgmental process could not be discounted, however, as other findings indicated that such assessment center measures as trait ratings and situational exercises nearly doubled the criterion variance explained (38%), compared to paper-and-pencil tests alone (20%). Thus, the assessment center was more useful when the assessment staff restricted its judgments to traits or exercises and did not integrate these and other data into a global rating. This issue of judgmental versus statistical integration of assessment center performance data requires further research. Hinrichs (1969) conducted a later study of the IBM program in which predictions were made for 47 lower level management candidates based on judgments made by an assessment center staff and judgments made by managers. Hinrichs noted that for this sample, managers using traditional methods were almost as

effective as the assessment staff. However, a validity coefficient of .37 was obtained using the assessment rating, while the traditional method obtained a correlation of .10. On the other hand, the interpersonal information generated by the assessment process was best estimated by the assessment center.

Recent IBM studies by Dodd (1971) and Kraut and Scott (1972) show significant findings based on assessment performance. Dodd, for example, summarized 11 studies conducted at IBM totaling 479 managers assessed. Correlations between assessment ratings and such criteria as position level, salary, change in position level and increase in salary ranged from .05 to .63 with a Mdn. r of .35. One interesting study reported by Dodd related assessment ratings and demotions. This study, because of its unequal criterion measure probably was less contaminated than most follow-up studies. Of those demoted due to job failure after receiving a promotion, only 4% of those seen as having high potential by the assessment center were demoted, while 20% of the lower potential group were subsequently demoted.

Kraut and Scott (1972) reviewed the career progress of 1086 employees promoted after attending an assessment center. Strong relationships between center performance and subsequent promotions, similar to the results obtained by A.T.&T. (Moses, 1971) were found. Of considerable interest was the follow-up of a small sample of individuals who were demoted after becoming first level managers. Of 167 candidates assessed, 14% were subsequently demoted. Significant differences between the demotion rates of high performing (4%) and low performing (20%) assessees were found. Since all demotions

occurred after the subjects had been promoted to management based on assessment data, and since demotions were based on job failure, these findings appear to be based on criterion data that is less contaminated than in most studies.

Myer (1972) was also able to study demotion rate. Of 36 foremen promoted without the GE assessment center, six were later removed for poor performance, while none of a somewhat matched group promoted partly based on assessment center information were demoted. In evaluating the results of this assessment center, General Electric has also found results similar to Campbell and Bray (1967) in that those assessed high in potential performed better on the job (1½ years later) than those assessed low in potential, but both groups did better than a third group that was not assessed at all. Because the groups were not matched it is impossible to tell if some kind of selection bias was working, or the results may be a reflection of the developmental effects of the assessment center.

Bentz (1971) presents a comprehensive review of validity data based on the Sears program in which a variety of rating criteria (ratings by personnel directors, superiors, self), as well as progress criteria (job mobility index and a salary progress index), were compared with various assessment predictors (tests, in baskets, group exercises). Sears has devoted considerable effort to validate each component of its assessment process and no single criterion measure is consistently used; therefore, direct comparisons are difficult to make. Considerable data, however, is reported which supports the relationship between the in basket, group simulations

and test performance data used in assessment and various indices of organizational success.

Bullard (1969) presents the only study showing that assessment centers were less valid than normal assessment procedures. This study, conducted at the Caterpillar Tractor Company, compared 37 individuals promoted after attending an assessment center with 27 individuals promoted based on traditional methods. None of the correlations between assessment ratings and job success ratings (which ranged from $-.09$ to $.26$) were statistically significant. The results are somewhat unclear, however, as only marginal individuals apparently were scheduled to attend the assessment center, while "obviously" superior candidates were promoted without assessment, thus raising a question about the representativeness of the samples studied. Another possibility is that the managerial demands of the "shop" supervisor position for which individuals were assessed, were not sufficiently critical, while the technical demands of the job accounted for most of the variance.

A study by Jaffee, Bender, and Calvert (1970) also compared a small sample ($N = 26$) of recently promoted individuals, half assessed and half promoted without assessment. Subordinates responded more favorably to those in the assessed group.

Another study by McConnel and Parker (1972) compared assessment ratings of 70 first-line supervisors to ratings of their present job performance. The assessment procedure successfully identified more than half of the job performance ratings.

Moses (in press) reports that a short assessment procedure resulted in comparable findings to a longer assessment process. Eighty-five men and women were assessed twice, once in an abbreviated assessment process and then in an expanded center. Considerable care was taken to prevent contamination of results. A correlation of .73 was obtained between the assessment performances in both programs. Of special interest is the fact that no differential validity results were found between any of the subgroups based upon race and sex and assessment performance. The only other study of differential validity reported was done by Meyer (1972). Of 120 whites assessed, 25% received offers while of 30 blacks assessed 50% received offers; thus indirectly a bias in favor of blacks was operating. However, such results are difficult to interpret, given the strong pressure on organizations to hire and promote blacks.

Another A.T.&T. study (Bray and Campbell, 1968) was unique in several ways. In addition to assessing sales rather than managerial skills, it also represented one of the few predictive validity studies reported. A job analysis of the salesman's job resulted in the development of a special assessment program designed to measure sales rather than supervisory management attributes or dimensions.

Several precautions were used to minimize criterion bias. In addition to the predictive validity format, where individuals were assigned to sales positions without reference to their assessment performance, a research assessment center was used, and special care was taken to obtain realistic criterion measures based on actual job performance. A team of sales reviewers was asked to independently

evaluate each subject after observing his performance in handling a variety of actual sales contracts. These ratings, as well as supervisor and trainer ratings, were obtained after each subject had been on the job for a minimum of six months.

The validity results obtained in this study were quite impressive. While the assessment center judged only half of the total group to be effective salesmen, all of those receiving a more than acceptable rating, and 68% of those seen as acceptable were also seen in the criterion evaluation as effective salesmen. Only 24% of the low rated assessees demonstrated adequate job performance. The correlation between the overall assessment rating and the field review rating was .51, a substantial increase in prediction compared with a multiple r of .33 between four paper-and-pencil tests and the criterion.

As impressive as these results were, the study pointed out the importance of using appropriate criterion measures. While the supervisor and trainer ratings had been gathered with considerable care, they were unrelated to the assessment center rating. These findings raise some serious methodological considerations where supervisors, trainers, or others rate job success of assessed groups for purposes of validity analyses. In these studies supervisors untrained in assessment have often been found to be biased in rating job success and in nominating men considered promotable.

As a general rule, about half of all individuals nominated as candidates to attend an assessment center are seen as not having sufficient abilities to succeed in management. Since the same supervisors who nominated candidates are often called on to evaluate

subsequent performance, the uncritical use of supervisory judgment as the sole basis of validity assessment performance is a questionable procedure. As noted by Bray and Moses (1972), it is somewhat absurd to validate an expensive, high powered (in the sense of information generated) process by a cheap, often unrealistic but easily obtainable "criterion" measure. Also, total reliance on supervisory nomination of candidates for assessment should also be questioned.

DISCUSSION AND SUGGESTED

TRENDS FOR RESEARCH

Unlike the mixed validity findings of the Initial Assessment Period, findings in the Research Assessment Period and Implementation Period have been consistently positive. To summarize the accuracy of prediction, as represented by the reported validity coefficients for the studies outlined in Table 1, median validity coefficients over studies were tabulated. Over all studies, the Mdn. r = .39; in predicting job performance, the Mdn. r = .32; in predicting job potential, the Mdn. r = .63; and in predicting job progress, the Mdn. r = .40. These findings do indicate that the global rating was robust, insofar as it was predictive of various criteria; especially job potential, a highly complex, multi-dimensional criterion. Particularly interesting is the question, what is the optimal time interval in which job potential will be predictable as manifest criterion behavior? The 2 to 8 years in which potential was manifested as progress (increases in salary, responsibility, promotions, etc.) in the validity studies in Table 1, might not be long enough, as the global rating remained more predictive of potential than progress.

The answer to the question can only be determined by additional longitudinal research, albeit costly, such as the Management Progress Study (Bray, 1964b).

The validity findings must also be interpreted with respect to the various design problems previously discussed, the most common problem being criterion contamination. Such bias, however, would seem to have greatest impact on such criteria as early progress or job performance, while the validity findings of the studies considered have been more positive in relationship to later progress and potential. Moreover, at least two research assessment center programs were designed with uncontaminated criteria, and positive validity findings were reported (Bray & Grant, 1966; Bray & Campbell, 1968). The Management Progress Study (Bray & Grant, 1966) was especially contamination free because the assessment staff consisted almost entirely of in-house and consultant psychologists and so the assessment staff could not even circumvent record keeping and leak feedback to line management.

If criterion contamination has resulted in inflated validity findings in some studies, there are probably several other factors having the opposite confounding effect of spuriously reducing validity findings. For example, the assessment center global rating has been highly robust as a predictor of various aspects of managerial success without any control over specific organizational conditions or specific job requirements. Predictive accuracy might substantially increase if these moderating conditions were identified and considered by the assessment staff in the assessment process.

Similarly, predictive accuracy might increase if selection into the assessment center was improved to be sure that all potential high performance people get into the center. Previously, top management probably has overlooked many high potential men among its assessment center nominees, as recent research studies at IBM have indicated (Dodd, 1970). Finally, the method by which the global rating was constructed also might affect its predictive accuracy. There is some evidence at IBM (Wollowick & McNamara, 1969) that the global rating constructed by the statistical integration of performance data is more efficient than the global rating constructed by judgmental integration, but only the latter rating procedure was used in all but one validity study.

Assessment center validity research should be extended beyond the above areas to study other factors which will improve the accuracy of prediction in the selection process, such as making differential validity analyses by race and sex; developing more effective methods of assessor training; establishing ways to control subject characteristics that reduce predictive accuracy; analyzing assessor rater biases; improving assessment exercises with more research devoted to determining what dimensions are best observed in each exercise; identifying the optimum number of dimensions for the assessment program and the best method to determine the dimensions; developing new ways for assessors to record behavior; and finally, developing new ways of evaluating and using assessment results.

Is the assessment center a more valid method than other traditional methods of assessment? When the traditional method was defined as

paper-and-pencil tests typically administered at the assessment center, the answer was unquestionably yes. The global rating explained 10% to 20% greater criterion variance than tests alone in several studies (Bray & Grant, 1966; Bray & Campbell, 1968; Carleton, 1970; Finley, 1970; Moses, 1971). In one study (Wollowick & McNamara, 1969), when test variance was partialled out of the global rating, the rating continued to explain 18% greater criterion variance.

When the traditional method was defined as sundry techniques used independently of the assessment center, a two-group design was used in which the subsequent rates of success of assessed and non-assessed groups considered promotable were compared. With the exception of one study (Bullard, 1969) in Table 1, the assessed groups were more successful. Median percentages of successes tabulated for both groups indicates the assessed group higher in subsequent job performance (.68/.59), job potential (.59/.28), and progress (.38/.08), so that the assessment center method again appears the more valid method. These results, however, cannot be interpreted conclusively because of various design problems, especially failure to adequately match assessed and non-assessed groups.

Much needed long-term cost savings data from the validity gain of using the assessment center method as compared to the traditional method is not available, although savings seem very substantial as the careers of successfully identified high potential men unfold after the assessment center. Even short-term cost savings data is hard to find. Dale Miller (1972) of Syntex made a study of the

various costs associated with hiring a poor sales manager in a position, rather than an average one. Figuring sales lost (or not gained), employee turnover, etc., he arrived at a conservative figure of \$50,000 a year. The results of this study suggest that even an assessment center whose predictive accuracy only moderately exceeds traditional assessment methods, has great management value.

Additional research also is necessary to evaluate the training value of the assessment center. The assessment center is strongly advocated as a means of increasing assessor skills in observing and recording behavior, performing selection and appraisal functions, and counseling individuals regarding performance. Assesseees are also thought to gain insight into group dynamics and career planning, and self-insight is said to be heightened; yet there are no published studies presently available to support these assumptions. The Meyer (1972) and Campbell and Bray (1967) articles suggest such learnings occur, but the basis and extent of the learnings is not clear. Performance measures before and after the assessment center must be compared for an assessed group and appropriate control groups, based upon such research designs as those suggested by McGehee and Thayer (1961).

Perhaps the severest criticism of assessment center research is its limited nature; that is, most research has concerned an operational A.T.&T. type assessment center developed for the selection of lower management candidates. Increased attention to validity extension (Mosier, 1951) is necessary, especially as the assessment center is used for the selection of non-managerial

applicants (i.e., black entrepreneurs, police officers, teachers), and new assessment center exercises, staffing practices, and administrative methods are developed.

Another subject for research is the question of the long-term organizational effects of introducing and enculturating an assessment center within an organization; i.e., development of new managerial climates, organizational adaptation to change, new organizational reward systems. This appears to be the most difficult area of research to pursue as it must be longitudinal, organization-wide, and well-controlled, whereas practicality markedly limits such research opportunities. Perhaps this research can proceed, however, in a controlled laboratory setting in which long-term mini-organizations are established.

CONCLUSIONS

The purpose of this article has been to review the literature concerning the criterion validity of assessment center programs in business and government organizations since 1956. The assessment staff global rating was focused upon as the predictor of subsequent job performance, potential, and progress. A secondary purpose of this article was to ascertain whether the assessment center or other traditional (psychometric) methods of assessment, such as paper-and-pencil tests, were more valid. The conclusions of this review are:

1. The assessment staff global rating consistently has been a valid predictor during the Research Assessment Period and Implementation Period. Predictive accuracy

varies with the criterion studied and is highest for job potential, followed by job progress, and then job performance.

2. The assessment staff global rating is robust as a predictor of various criteria. However, assessment center programs have been limited to the assessment of management and sales groups, and so the extent of robustness of the global rating has not been fully explored.
3. Assessment center validity research has been subject to various design problems which probably have both suppressed and inflated validity findings. Although few studies have had sufficient controls to avoid these problems, those that have also have demonstrated positive validities as high or higher than studies with design problems.
4. The assessment center clearly is a more valid method than paper-and-pencil tests. Much evidence also suggests that it is a more valid method than other traditional methods of assessment in terms of subsequent rates of success of assessed and non-assessed groups.
5. Little effort has been made to identify and control those factors moderating the relationship between the assessment staff global rating and the criterion.
6. Little effort has been made to evaluate the usefulness of the assessment center as a training tool or as an organizational development approach.

References

- Barron, F. Personal soundness in university graduate students.
Berkeley:University of California Press, 1954.
- Bentz, V. J. The Sears experience in the investigation, description and prediction of executive behavior. In F. R. Wickert & D. E. McFarland (Eds.), Measuring executive performance. New York: Appleton-Century-Crofts, 1967. Pp. 147-205.
- Bentz, V. J. Validity studies at Sears. In W. C. Byham (Chm.), Validities of assessment centers. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., 1971.
- Bray, D. W. The assessment center method of appraising management potential. In J. W. Blood (Ed.), The Personnel Job in a Changing World. New York: American Management Association, 1964. (a)
- Bray, D. W. The management progress study. American Psychologist, 1964, 19, 419-420. (b)
- Bray, D. W., & Campbell, R. J. Selection of salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.
- Bray, D. W., & Grant, D. L. The assessment center in the measurement of potential for business management. Psychological Monographs, 1966, 80, (17, Whole No. 625).
- Bray, D. W., & Moses, J. L. Personnel selection. Annual Review of Psychology, 1972, 23, 545-576.
- Bullard, J. F. An evaluation of the assessment center approach to selecting supervisors. Mimeo. Report, Caterpillar Tractor Co., Corporate Personnel Department, May 1969.

Byham, W. C. Assessment center for spotting future managers. Harvard Business Review, 1970, 48 (4), 150-160, plus appendix.

Byham, W. C. The assessment center as an aid in management development. Training and Development Journal, Dec. 1971, 25, 10-22.

Byham, W. C. A new way to identify management potential. European Business, in press.

Campbell, R. J., & Bray, D. W. Assessment centers: An aid in management selection. Personnel Administration, 1967, 30 (2), 6-13.

Carleton, F. O. Relationships between follow-up evaluations and information developed in a management assessment center. Paper presented at the meeting of the American Psychological Association, Miami Beach, Florida, 1970.

Chronbach, L. J. Essentials of psychological testing. New York: Harper & Row, 1970.

Dodd, W. E. Will management assessment centers insure selection of the same old types? Paper presented at the meeting of the American Psychological Association, Miami Beach, Florida, 1970.

Dodd, W. E. Validity studies at IBM. In W. C. Byham (Chm.), Validity of assessment centers. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., 1971.

Dunnette, M. D. Multiple assessment procedures in identifying and developing managerial talent. In P. McReynolds (Ed.), Advances in psychological assessment, Vol. II. Palo Alto: Science & Behavior Books, 1971.

Finkle, R. B., & Jones, W. S., Assessing corporate talent. New York: Wiley, 1970.

Finley, R. M., Jr. An evaluation of behavior predictions from projective tests given in a management assessment center. Paper presented at the meeting of the American Psychological Association, Miami Beach, Florida, 1970.

General Electric Company. Talent Development Program (Supplement for Potential Staff Members). Mimeo. Report, Corporate Personnel Department, 1969.

Gough, H. G. The construction of a personality scale to predict scholastic achievement. Journal of Applied Psychology, 1953, 37, 361-366.

Grant, D. L., & Bray, D. W. Contributions of the interview to assessment of management potential. Journal of Applied Psychology, 1969, 53, 24-34.

Grant, D. L., Katkovsky, W., & Bray, D. W. Contributions of projective techniques to assessment of management potential. Journal of Applied Psychology, 1967, 51, 226-232.

Greenwood, J. M., & McNamara, W. J. Interrater reliability in situational tests. Journal of Applied Psychology, 1967, 31, 101-106.

Harriss, H. The group approach to leadership testing. London: Kegan Paul, 1959.

Hinrichs, J. R. Comparism of "Real Life" assessment of management potential with situational exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology, 1969, 53, 425-433.

Holmen, M. G. An assessment program of OCS applicants. HumRRO Technical Report No. 26, 1956.

Holt, R. R. & Luborsky, L. Personality patterns of psychiatrists.

New York: Basic Books, 1958.

Jaffee, C. L. Managerial assessment: Professional or managerial prerogative. Personnel, 1966, 45, 162-163.

Jaffee, C. L. A tridimensional approach to management selection. Personnel, 1967, 46, 453-455.

Jaffee, C., Bender, J., & Calvert, D. L. The assessment center technique: A validation study. Management of Personnel Quarterly, Fall 1970, 9, 9-14.

Kelly, E. L., & Fiske, D. W. The prediction of performance in clinical psychology. Ann Arbor: University of Michigan Press, 1951.

Kraut, A. I., & Scott, G. J. Validity of an operational management assessment program. Journal of Applied Psychology, 1972, 56, 124-129.

MacKinnon, D. W. An assessment study of Air Force officers. Part V: Summary and applications. WADC Technical Report 58-91 (V), Wright Air Development Center, 1958.

Mather, B. E. Assessment of men: A validity study. Unpublished masters thesis, Massachusetts Institute of Technology, 1964.

McConnel, J., & Parker, T. An assessment center program for multi-organizational use. Training and Development Journal, March 1972, 26, 6-14.

McGehee, W., & Thayer, P. W. Training in business and industry. New York: Wiley, 1961.

Meyer, H. H. The validity of the in-basket test as a measure of managerial performance. Personnel Psychology, 1970, 23, 297-307.

- Meyer, H. H. Assessment centers at General Electric. Paper presented at the meeting of Development Dimensions, San Francisco, Calif., 1972.
- Miller, D. Assessment centers at Syntex. Paper presented at the meeting of the Personnel Division of the National Pharmaceutical Assoc., Hot Springs, Va., 1972.
- Moore, B. Personnel assessment centers in Britain and America. Mimeo. Report, Sidney, Australia: New South Wales Public Service Board, December 1969.
- Morris, B. S. Officer selection in the British Army, 1942-1945. Occupational Psychology, 1949, 23, 219-234.
- Mosier, C. I. Problems and designs in cross-validation. Educational and Psychological Measurement, 1951, 11, 5-11.
- Moses, J. L. On the application of clinical method in personnel psychology. Paper presented at the meeting of the New York State Psychological Association, Kiamisha Lake, N. Y., May 1969.
- Moses, J. L. Assessment center performance and management progress. In W. C. Byham (Chm.), Validity of assessment centers. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., 1971.
- Moses, J. L. The development of an assessment center for the early identification of supervisory potential. Personnel Psychology, In press.
- Murray, H. Explorations in personality. Cambridge: Oxford University Press, 1938.
- OSS Assessment Staff. Assessment of men. New York: Rinehart, 1948.

Stern, G. G., Stein, M. I., & Bloom, B. S. Methods in personality assessment. Glencoe, Ill.: Free Press, 1956.

Taft, R. Multiple methods of personality assessment. Psychological Bulletin, 1959, 56, 333-352.

Thomson, H. A. Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.

Vernon, P. E. The validation of Civil Service Selection Board Procedures. Occupational Psychology, 1950, 24, 75-95.

Wollowick, H. B., & McNamara, W. J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.

Footnote

¹ Requests for reprints should be sent to Barry M. Cohen, Faculty of Management and Psychology, University of West Florida, Pensacola, Florida 32504.

TABLE I

A SUMMARY OF RESEARCH STUDIES

Study	Main Objective	Location	Assessors	Groups Studied	N	Criteria
Bray & Grant, 1966	research program, Management Progress Study	A T & T	psychologists & a few managers	young managers - college sample - non college sample	123 144	progress in management
Bray, 1964a	operational program	A T & T	managers	recently promoted assessed and non-assessed managers	80	job performance potential for advancement
Mather, 1964	operational program	A T & T	managers	recently promoted assessed managers	42	job performance
Campbell & Bray, 1967	operational program	A T & T	managers	recently promoted assessed and non-assessed managers	501	job performance potential for advancement
Bray & Campbell, 1968	research program	A T & T	sales managers	recently employed salesmen	78	field review of performance
Vollovich & McNamara, 1969	operational program	IBM	managers	managerial candidates	94	increase in managerial responsibility after 3 year
Hinrichs, 1969	operational program	IBM	managers	managerial candidates	47	salary
Bullard, 1969	operational program	Caterpillar Tractor Co.	general foremen	assessed and non-assessed operatives recently promoted to general foremen	64	job performance ratings

TABLE 1 (Cont.)

Study	Main Objective	Location	Assessors	Groups Studied	N	Criteria
Carleton, 1970	operational program	Standard Oil (Ohio)	psychologists & managers	managerial candidates	122	ratings a. 12 traits b. job potential c. salary & promotion procedure
Finley, 1970	operational program	Standard Oil (Ohio)	psychologists & managers	managerial candidates	109	ratings a. 122 traits b. job potential
Jaffee, et al., 1970	operational program	ORNL, Nuclear Div.	managers	recently promoted managers	26	feelings of supervisor and subordinates
Thomson, 1970	operational program	Standard Oil (Ohio)	psychologists & managers	management candidates	71	supervisory ratings
Dodd, 1971	operational program	IBM	managers	management candidates	479	change in management level, change in salary
Bentz, 1971	operational program	Sears	psychologists & managers	managerial trainees	100	job performance ratings & progress in management
Moses, 1971	operational program	A T & T	managers	management candidates	5,943	progress in management
Kraut and Scott, 1972	operational program	IBM	managers	management candidates	1,086	progress in management
McConnel and Parker, 1972	operational program	three manufacturing organizations and a bank	managers	first-level supervisors	70	ratings of present job performance
Notes, in press	research	A T & T	managers	short service	85	assessment

ASSESSMENT CENTER
BIBLIOGRAPHY

- 1) Acker, S. R. & Perison, M. R. Can we sharpen our management of human resources? Corporate Personnel Department, Company Distribution, Olin Corporation, June, 1970.
- 2) Albrecht, P. A., Glaser, E. M., & Marks, J. Validation of a multiple assessment procedure for managerial personnel. Journal of Applied Psychology, 1964, 43, 351-359.
- 3) Albrook, R. C. How to spot executives early. Fortune, July 1968, 78 (2), 106-111.
- 4) Assessment of Men. OSS Assessment Staff, New York: Rinehart, 1948.
- 5) Bass, B. M. The leaderless group discussion. Personnel Psychology, 1954, 7, 470-477.
- 6) Bender, J. M., Calvert, O. L., and Jaffee, C. L. Report on supervision selection program Oak Ridge Gaseous Diffusion Plant, April 17, 1970, (K 1789), Union Carbide Corporation, Nuclear Division.
- 7) Bentz, V. J. The Sears experience in the investigation, description, and prediction of executive behavior, Executive Study Conference, Educational Testing Service, 1962.
- 8) Bentz, V. J. Validity of Sears assessment center procedures. Paper presented as part of the symposium, "Validity of Assessment Centers", at the 79th Annual Convention of the American Psychological Association, 1971.
- 9) Bray, D. W. The management progress study. American Psychologist, 1964, 19, 419-420.
- 10) Bray, D. W. The assessment center method of appraising management potential. In J. W. Blood (Ed.), The Personnel Job in a Changing World, New York: American Management Association, 1964.
- 11) Bray, D. W. & Grant, D. L. The assessment center in the measurement of potential for business management. Psychological Monographs, 1966, 80 (17, Whole No. 625).
- 12) Bray, D. W. & Campbell, R. J. Selection of salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.
- 13) Bray, D. W. Assessment Centers. Unpublished paper given at New York Workshop on Assessment Centers, 1969.
- 14) Bray, D. W. In-basket as used in the assessment centers. Paper presented at the New York Workshop on Assessment Centers, New York, 1969.
- 15) Bray, D. W. The uses of assessment center methods. Projector, April 1971, 1-3.

- 16) Bray, D. W. Increased opportunities for women through management assessment center, Personnel, September-October 1971, 31-34.
- 17) Bray, D. W., & Moses, J. L., Personnel selection. Annual Review of Psychology, 1972, 23, 545-576.
- 18) Bullard, J. F. An Evaluation of the assessment center approach to selecting supervisors. Mimeo report, May, 1969, Caterpillar Tractor Company.
- 19) Byham, W. C. The uses of assessment centers. Paper presented at Spring 1969 meeting of the Executive Study Conference, 1969.
- 20) Byham, W. C., & Pentecost, R. The assessment center: Identifying tomorrow's managers. Personnel Magazine, September-October, 1970.
- 21) Byham, W. C. Assessment center for spotting future managers. Harvard Business Review, 1970, 48 (4), 150-160, plus appendix.
- 22) Byham, W. C., & Thornton, G. C., III. Assessment centers: new aid in management selection. Studies in Personnel Psychology, 1970, 2 (2), 21-35.
- 23) Byham, W. C. The assessment center as an aid in management development. Training and Development Journal, December 1971.
- 24) Byham, W. C. A new way to identify management potential. European Business, in press.
- 25) Campbell, J. T., Otis, J. L., Liske, R. E., & Prien, E. P. Assessments of higher-level personnel: II, Validity of the overall assessment process. Personnel Psychology, 1962, 15, 63-74.
- 26) Campbell, R. J., & Bray, D. W. Assessment Centers: an aid in management selection. Personnel Administration, 1967, 30, (2), 6-13.
- 27) Carleton, F. O. Relationships between follow-up evaluations and information developed in a management assessment center. Paper presented at the meeting of the American Psychological Association Convention, Florida, 1970.
- 28) Churchill, G. Joblot: A production management game, New York: Macmillan, 1970.
- 29) Cronbach, L.J. Essentials of psychological testing. (3rd Ed.), Harper & Row, 1970.
- 30) Cronbach, L. J. Essentials of psychological testing. (2nd Ed.), Harper & Row, 1960.

- (31) Crooks, L. A. Issues in the development and validation of in-basket exercises for specific objectives. Research Memorandum 68-23. Princeton, N. J.: Educational Testing Service, 1968. Paper presented at the meetings of the American Psychological Association, Symposium on the "In-Basket Exercise--Test or Technique?" San Francisco, September, 1968.
- (32) DiCostanzo, F. & Andretta, T. The supervisory assessment center in the Internal Revenue Service. Training and Development Journal, 1970, 12-15.
- (33) Dodd, W. E., & Kraut, A. I. Will management assessment centers insure selection of the same old types? Paper presented at the meetings in the American Psychological Association Convention, Florida, 1970.
- (34) Dodd, W. E., & Kraut, A. I. The prediction of management assessment center performance from earlier measures of personality and sales training performance. A preliminary report, April 1970, Corporate Personnel Research.
- (35) Dodd, W. E. Two factors related to career progress in IBM: A summary of longitudinal studies in the identification of management potential.
- (36) Dodd, W. E. Summary of IBM assessment validations. Paper presented at the symposium, "Validity of Assessment Centers", at the 79th Annual Convention of the American Psychological Association, 1971.
- (37) Donaldson, R. J. Validation of the internal characteristics of an industrial assessment center program using the multitrait-multimethod matrix approach. Unpublished dissertation, Case Western Reserve University, 1969.
- (38) Dunnette, M. D. Multiple assessment procedures in identifying and developing managerial talent. In P. McReynolds (Ed.), Advances in Psychological Assessment, Vol. II. Palo Alto: Science & Behavior Books, 1971.
- (39) Educational Testing Service. The in-basket technique. Conference on the Executive Study. Princeton, New Jersey: ETS, 1960.
- (40) Educational Testing Service. Management games in selection and development. Conference on the Executive Study. Princeton, New Jersey: ETS, 1964.
- (41) Eeny, meeny, miney . . . supervisor? Industry Week, November 16, 1970, 28-32.
- (42) Finkle, R. D., & Jones, W. S. Assessing Corporate Talent: a key to managerial manpower planning. New York: Wiley-Interscience, 1970.

- (43) Finley, R. M., Jr. An evaluation of behavior predictions from projective tests given in a management assessment center. Paper presented at the meetings of the American Psychological Association Convention, Florida, 1970.
- (44) Flanagan, S. C. "Some Considerations in the Development of Situational Tests." Personnel Psychology, 1954, 7, 461-464.
- (45) Frederiksen, N. Factors in in-basket performance. Psychological Monographs, 1962, 76 (22, Whole No. 541).
- (46) Frederiksen, N., Saunders, D. R., & Wand, B. The In-basket Test. Psychological Monographs, 1957, 71 (9, Whole No. 438).
- (47) General Electric, Talent Development Program (Suppliment for Potential Staff Members) Mimio Report, 1969.
- (48) Ghiselli, E. E. Managerial talent. American Psychologist, 1963, 8, 631-642.
- (49) Gordon, L. V. Clinical, psychometric, and work-sample approaches in the predection of success in Peace Corps Training. Journal of Applied Psychology, 1957, 51, 111-119.
- (50) Grant, D. L., and Bray, D. W. Contributions of the interview to assessment of management potential. Journal of Applied Psychology, 1969, 53, 24-34.
- (51) Grant, D. L., Katkovsky, W. and Bray, D. W. Contributions of projective techniques to assessment of management potential.. Journal of Applied Psychology, 1967, 51, 226-232.
- (52) Greenwood, J. M. & McNamara, W. J. Leadership styles of structure and consideration and managerial success. Personnel Psychology, 1968, 21, 85-98.
- (53) Greenwood, J. M. & McNamara, W. J. Interrater reliability in situational tests. Journal of Applied Psychology, 1967, 31, 101-106.
- (54) Guion, R. M. Personnel Testing. New York: McGraw-Hill, 1965.
- (55) Guyton, T. "The Identification of executive potential." Personnel Journal, 1969, 48 (11), 866-872.
- (56) Hardesty, D. L. & Jones, W. S. Characteristics of judged high potential management personnel - The operations of an industrial assessment center. Personnel Psychology, 1968, 21, 85-98.
- (57) Harriss, H. The Group Approach to Leadership Testing. London: Kegan Paul, 1959.

- (58) Hinrichs, J. R. Comparison of "Real Life" assessments of management potential with situational exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology, 1969, 53, 425-433.
- (59) Holman, M. G., Katter, R. V., Jones, A. M., and Richardson, I. F. An assessment program for OCS applicants. Humro Technical Report, 26, 1956.
- (60) Jaffee, C. L. Assessment centers help find management potential. Bell Telephone Magazine, 1965, 44 (3), 18-24.
- (61) Jaffee, C. L. Managerial assessment: Professional or managerial prerogative. Personnel Journal, 1966, 45, 162-163.
- (62) Jaffee, C. L. A tridimensional approach to management selection. Personnel, 1967, 46, 453-455.
- (63) Jaffee, C. L., Bender, J., & Calvert, D. The assessment center technique: a validation study. Management of Personnel Quarterly, Fall, 1970, 9-14.
- (64) Kohles, J. "Evaluation of the assessment center approach to the selection of college recruits in the eastern territory." International Report. Chicago Sears and Robuck Co., Psychological Research and Service Section, 1970.
- (65) Korman, A. K. The prediction of managerial performance. Personnel, 1969, March/April.
- (66) Krout, A. I., and Scott, G. S. Validity of an operational management assessment program. Journal of Applied Psychology, April 1972, 56 (2), 124-129.
- (67) Laurent, H. A. "Summary of Early Identification of Management Potential Research Project", 1961, Social Science Research Division of the Employee Relations Department, Standard Oil Company.
- (68) Lindert, P. H. MACRO: A Game of Growth and Policy. New York: Holt, Rinehart and Winston, 1970.
- (69) Mather, B. E. Assessment of men: A validity study. Unpublished masters thesis, Massachusetts Institute of Technology, 1964.
- (70) McConnell, J. H. The assessment center in the smaller company. Personnel, 1969, March/April.
- (71) McConnell, J. H. The assessment center: a flexible program for supervisors. Personnel, September/October, 1971.

- (72) McFarlan, F. The Management Game, New York: Macmillan, 1970.
- (73) McGehee, W. & Thayer, P. W. Training in Business and Industry. New York: Wiley, 1961.
- (74) McKinnon, D. W. An assessment study of Air Force Officers. Technical Report WADC - TR - 59 - 91. Washington D. C. Office of Technical Services, United States Department of Commerce, 1958.
- (75) McNamera, W. J., and Greenwood, S. M. Evaluation of the personnel assessment program. Personnel Research Studies, IBM, 1967.
- (76) Meyer, H. H. The validity of the In-basket test as a measure of managerial performance. Personnel Psychology, 1970, 23 (3), 297-307.
- (77) Meyer, H. H. Assessment centers at General Electric. Paper presented at the meeting of Development Dimensions, San Francisco, Calif., 1972.
- (78) Miller, D. Assessment centers at Syntex. Paper presented at the meeting of the Personnel Division of the National Pharmaceutical Assoc., Hot Springs, Va., 1972.
- (79) Moor, B. Personnel assessment centers in Britain and America. Mimio Report, Sydney, Australia: New South Wales Public Service Board, December, 1969
- (80) Morris, B. S. Officer-selection in the British Army, 1942-1945. Occupational Psychology, 1949, 23, 219-234.
- (81) Moses, J. L. On the application of clinical method in personnel psychology. Paper presented at the meeting of the New York State Psychological Association, Kiamisha Lake, N. Y., May, 1969.
- (82) Moses, J. L. The development of an assessment center for the early identification of supervisory potential. Personnel Psychology, in press.
- (83) Moses, J. L. Assessment center performance and management progress. Studies in Personnel Psychology, 1972, 4, 7-12.
- (84) Moses, J. L. Assessment center performance and management progress (AT&T). Paper presented as part of the symposium, Validity of Assessment Centers, at the 79th Annual Convention of the American Psychological Association, 1971.
- (85) Murray, H. Explorations in Personality. Cambridge: Oxford University Press, 1938.

- (86) Prien, E. D. Assessments of higher level personnel: an analysis of interviewers' predictions of job performance. Personnel Psychology, 1962, 15, 319-344.
- (87) Sadoka, J. M. Factor analysis of OSS situation tests. Journal of Abnormal & Social Psychology, 1962, 47, 843-852.
- (88) Schaffer, A. J. Information about assessment center for ES&D Program Finalists, Memorandum from Director, Personnel Division, National Office, Internal Revenue Service, September 16, 1970.
- (89) Stern, G. G., Stein, M. I., & Bloom B. S. Methods in personality assessment. Glencoe, Ill., Free Press, 1956.
- (90) Story, W. Assessment Centers. Paper presented at New York Workshop on Assessment Center, New York, 1969.
- (91) Taft, R. Multiple methods of personality assessment. Psychological Bulletin, 1959, 56, 333-352.
- (92) Talent Development Program (Supplement for Potential Staff Members) Mimeo Report, 1969, General Electric.
- (93) Taylor, E. K., & Nevis, E. C. The use of projective techniques in management selection. Personnel, 1952, 33, 462-474.
- (94) Tennessee Valley Authority, TVA's experiment in the assessment of managerial potential. Internal publication, undated.
- (95) Thomson, H. A. Internal and external validation of an industrial assessment program. Unpublished doctoral dissertation, Case Western Reserve University, 1969.
- (96) Thomson, H. A. Comparison of predictor and in criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.
- (97) Thornton, G. C. Paper presented at the symposium, Validity of Assessment Centers, at the 79th Annual Convention of the American Psychological Association, 1971.
- (98) Turney, J. R. The early identification of managerial potential in a technical-professional organization. Proceedings of the Annual Convention of the American Psychological Association, 1971, Vol. 6 (Pt. 2), 481-482.
- (99) Vernon, P. E. The validation of civil service selection board procedures. Occupational Psychology, 1950, 24, 75-95.

- (100) Where they make believe they're the boss. Business Week, August 28, 1971, 34-35.
- (101) Whisler, T. L., & Harper, S. F. Performance appraisal: Research and practice. Holt, Rinehart & Winston, 1962.
- (102) Wollowick, H. B. & McNamara, W. J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.

11-11

THE USE OF DIAGNOSTIC SMALL GROUPS IN POLICE RECRUIT SELECTION

1. Background. Dr. Robert B. Mills, Professor of Psychology at the University of Cincinnati's Department of Criminal Justice, formulated and now heads Cincinnati's method for evaluating the psychological aptitude of police candidates. This psychological evaluation is the final step in a screening procedure which includes an AGCT intelligence test, physical examination, individual interview, polygraph and background investigation. This method has aided in selection of the majority of the present Cincinnati police force.¹

2. Description of the Evaluation.

a. The evaluation process requires about five hours and involves candidates in small diagnostic groups of 8-12 men. The first step is the application of a combination of objective and projective-type personality tests administered along with several situational tasks. The philosophy here being that a combination of approaches offers more information than a single approach, and that the careful diagnostic integration of several different sources of information on each candidate by a control team represents the most equitable and thorough evaluation.²

1. How to Pick a Pro, Cincinnati Horizons, the Magazine of the University of Cincinnati, Dec 73.

2. Personal Communication, Dr. Robert Mills, Jan 74.

b. The basic evaluation or test consists of three parts. A Foot Patrol Observation (a walk of six blocks through a black business district followed by 25 multiple choice questions and an essay). A Clues Test (a ten-minute investigation of a hypothetical business employee's disappearance followed by a questionnaire dealing with both facts and hypothesis), and a Bull Session (a two-hour informal group discussion period). The evaluation sheet used for rating purposes by the four control personnel (two psychologists and two psychology graduate students) is shown at Inclosure 1.

c. The cornerstone of the psychological evaluation is the diagnostic group session termed "The Bull Session."³ This is the final two-hour wrap-up in which the team confronts 8-12 candidates who have completed all psychological tests. Members of the evaluation team have been briefed on the strengths and weaknesses of each candidate through an analysis of tests previously administered. At this point the evaluators are ready to witness the leadership styles and personal qualities of the candidates. Two evaluators are designated as group leaders while the remaining two members serve as participants-observers. Group leaders assemble the candidates around a conference table, inform them that their participation is essential, and that failure to make views and attitudes known during the session could result in a poor recommendation. The latter is a precaution against attempting to hide within the group.

3. Personal Communication, Dr. Robert Mills, Jan 74.

d. A number of exercises derived from real life police situations are used to allow the applicant a chance to step into the role of a policeman and cope with a typical stress situation that might confront any officer on duty. Each applicant gets a chance to play policeman while the other applicants and evaluators analyze his performance. A list of typical situations used in the Bull Session is at Inclosure 2.

e. The following is a detailed description on an exercise experienced during the period. This method was actually reviewed as part of the crimes of violence psychological testing (3F) project:

"You've been called to a carry-out restaurant on Reading Road," Dr. Mills had explained. "It's a Friday night. It's crowded. The manager wants you to remove a young black guy from the parking lot. Why? Because the manager says he called one of the waitresses a hill-billy whore. The manager doesn't want to sign a formal complaint. He just wants the man removed. Got it?"

"Got it," the lanky blond boy had replied.

At last, the slow-moving "cop" could saunter no further. He was standing alongside the metal folding chair, the "car," and gazing down at the still "cool" youth who was sprawled out in a leisurely pose, purposefully ignoring "the man" at his side.

Hooking his right thumb on his belt, the "cop" raised his gaze toward the nonexistent sky, the ceiling of the room. A look of confidence settled on his face. And he stood there for what might have been a minute, poised and posed like a peacock. Then, in an unmistakably derisive tone of voice, he spit his words out of the corner of his mouth, as if he were ridding himself of some vile-tasting tobacco juice. "Get your butt out of that car before I haul it out myself."

There was silence.

"All right, I said get your butt out of that car!" the blond boy persisted. "Or do you want me to pull you out?"

Still leaning back in his chair, the young black man moved the toothpick from one side of his mouth to the other. Calmly, he replied, "I know what they told you. But man, I didn't call that white broad no whore. She ain't got enough sense to be a whore. All she knows how to do is fry hamburgers."

Laughter rang out from around the circle of observers. But the "cop" paid no attention. "You gonna do what I ask, kid?"

"You got any brains, cop?"

At that, the blond boy grabbed the black one, pulling him to his feet by clutching the front of his shirt. But what to do next? What to say? Standing there, in a classically offensive pose, he seemed to freeze. His mind raced past a hundred possible plans of action, embracing none of them, while he continued standing there, rather ridiculously, clutching the "cool" black youth as if he were holding a dog by the scruff of the neck.

Suddenly a voice came from the crowd, "Hey, look at that honky pig shakin' down the brother!"

A reply rang out. "Yeah man, a dumb lookin' dude, ain't he? Brave though, and tough."

"Wish I was a cop," whined another voice.

Confused, the blond boy let loose of his captive. He glanced quizzically toward the psychologists, turned, and headed back to his chair. "I thought there were only two of us involved," he feebly protested.

Dr. Evangeline Norton, U.C. psychologist, and one of the voices in the crowd, answered his remark. "You think that on a Friday night, at a favorite hangout, there's going to be just the two of you involved?"

"Guess not," replied the youth.

"That scene," says Bob Mills, "seems to be consistently badly done. Yet it treats of almost all the issues we like to consider."

"If common sense predominates, the officer will handle the situation somewhat gingerly. First of all, he's operating under hearsay evidence. But what's more, he's one white officer in the midst of a throng of black juveniles, who, for the most part, have very negative attitudes toward police. Often, by the time a candidate has finished his role-playing, World War III is about to commence. It's a classical blunder that could only be made by an untrained or otherwise unsatisfactory police officer."

f. The above situation, and others like it, all contribute to spotting potentially unsatisfactory police officers. It takes four to six hours for Dr. Mills and his assistant to evaluate the entire group of applicants. Dr. Mills has developed the following guidelines in his search for a definition of just what qualities are to be looked for in a policeman. Mills calls his guidelines or desired qualities "the seven deadly virtues" of a good policeman.⁴ These virtues represent personality traits he feels will provide a policeman with survival value.

4. How to Pick a Pro, Cincinnati Horizon, Dec 73.

(1) Motivation for a law enforcement career. A reasonable and objective rationale which includes consideration of risks and benefits of a law enforcement career.

(2) Normal masculine self assertion. The ability to take charge of an emergency situation. Although training is a key element, the roots of that ability are in the psychological makeup of the officer.

(3) Emotional stability and good judgment under stress. Self-explanatory. The role playing of the candidate is ideally suited for this evaluation.

(4) Sensitivity toward minority groups and social deviates. An open attitude and understanding of various minority groups is rare. For example, white applicants from the suburbs are astonishingly naive about sexual deviates, alcoholics, prostitution, and racial minorities.⁵ If the applicant seems open in his attitude, he is considered trainable. If he is close-minded and ignorant about such areas, there is strong indication his impartiality as a police officer is jeopardized. The lack of impartiality should be grounds for disqualification.

5. Personal Communication, Dr. Robert Mills, Jan 74.

(5) Collaborative leadership skills. The degree to which the applicant is able to express his ideas and self to include his ability to be socially outgoing and influence others. All of these factors are measures of his potential success in law enforcement.

(6) Flexibility. The new breed of law enforcement officers are called upon to handle a wide and diverse range of situations all in a single duty day. His duty brings him in contact with community meetings, domestic disturbances, gathering and presenting evidence, school authorities, juvenile addicts, parents, etc., all requiring different approaches.

(7) Mature relationship to social authority. A man who is capable of exerting independent judgment in constructive ways. Screening for this quality precludes the selection of applicants with troublesome personalities such as the passive aggressive.

Dr. Mills admits that finding a person who possesses an overabundance of all seven characteristics may be an impossibility. But he feels that strengths in certain areas counterbalance weaknesses in other areas.⁶ His evaluation techniques look at the applicant as a whole person. In the evaluation the applicant must evaluate his role playing performance. In fact, the applicant gets the first opportunity at the critique.

6. Personal Communication, Dr. Mills, Jan 74.

3. Evaluation.

a. The results of Dr. Mills' first effort commencing in 1966 indicated that situational tests were not correlated with police academy grades. A significant correlation of .38 was found between the Clues Test and Police Academy standing based on a combination of final rank and a pistol test score. A significant correlation of .60 was found between academy standing and the Army General Classification Test. Of interest was the fact that the Clues Test and AGCT scores were not significantly correlated with each other (.105). In this case the Clues Test may measure behavior dimensions not represented in paper and paper intelligence tests. These behavior dimensions later known as the seven virtues may be important performance criteria. Further, they seem to be independent of those aspects measured by the Army General Classification Test.

b. The constant revision and improving of this program since 1966 has resulted in a standardized system. Captain Norman Hughes, Police Personnel Officer, Cincinnati Police Department, expressed the opinion that Mills' psychological evaluation is one of the most important facets of the entire police applicant examination processes.⁷ It was obvious to all contacted during the Mills method study that as a selection technique it is both valuable and accurate. But as is any selection process, one aspect alone

7. Personnel Communication, Captain Norman Hughes, Jan 74.

cannot produce the desired results. The combination of tested techniques is required to produce the optimum. The City of Cincinnati, like the US Army, uses several screening techniques or standards in the selection of their law enforcement personnel.

c. In my discussion with Dr. Mills, he indicated that it was possible that his method could readily be adapted for use as a screening program for US Army law enforcement personnel. Since the psychological evaluation phase should be completed prior to commencing the AIT or Basic Law Enforcement Course phase, the cost in establishing teams at each Basic Training Center would be extensive and not recommended as practicable at present. However, this system could be implemented in conjunction with the new One-Station Training System at the US Army Military Police School NLT FY 77 at a cost of 1 man-day's effort per 15 soldiers and a total of 10 spaces (3 civilian and 7 military). This would provide an evaluation program chief and three teams capable of conducting 15 sessions per week (5 per team), 15 soldiers per session.

CONFIDENTIAL - For Professional Use Only

PSYCHOLOGICAL EVALUATION
Police Cadet/Recruit

Candidate	Age	Ethnic Identity
-----------	-----	-----------------

Present

Marital Status	Education	Employment
Single	High School	Unemployed
Married	College	Employed
Divorced	High School	Unemployed
Widowed	College	Employed

Employment

**Overall
Rating**

Not Obs.	High Risk	Marginal	Average	Above Aver.	Superior
	0	1	2	3	4

I. INTELLECTUAL EFFECTIVENESS

Considers relevant factors before taking action

Shows awareness of future implications of decisions

**Decisive; able to make decisions
after considering alternatives**

Readiness and eagerness for new learning

Overall intellectual ability

II. PERSONALITY CHARACTERISTICS

Composure under stress; under- or over-reaction

Appropriate use of authority in relation to peers

App. use auth. in rel. to citizens

App. use auth. in rel. to subordinates

App. use auth. in rel. to superiors

Appropriateness of anxiety

Appropriateness of defensiveness

Adaptability; flexibility

Forcefulness; approp. aggressiveness

Assertive without alienating others

Willing to take a stand; risk-taking

Able to influence others with variety
of persuasive techniques

Awareness of ethnic and social diversity

Tolerance of ethnic and social diversity

Competence to deal with ethnic and social diversity

Sensitivity to needs of own ethnic group

IV. INTERPERSONAL RELATIONSHIPS

Self-awareness of personal impact

Ability to accept feedback about own behavior

Ability to give feedback and accurately assess others
Acceptance by peers

V. MOTIVATION FOR POLICE CAREER

VI. CONTRIBUTORY PREVIOUS EXPERIENCE

VII. COMMUNITY SERVICE ORIENTATION

VIII. TESTS, INTERVIEW, OR OTHER EVIDENCE

[illegible]

Descriptive Summary:

Rated by:

Date of evaluation:

POLICE RECRUIT/CADET ROLE-PLAYING EXERCISE
(LET'S SEE YOU TALK YOUR WAY OUT OF THIS ONE)

Situation 1: A break-in of a Pony Keg in Oakley has been reported by two juveniles. While cruising the area, you intercept a juvenile some time later staggering down a back street. He appears intoxicated, and you recognize him as the same juvenile implicated in previous break-ins in the neighborhood. Your initial interrogation proves fruitless; the juvenile admits nothing, and refuses to name his associate. You place him under arrest, and take him to the district for further questioning. Your sergeant says, "Yes, I know this punk. Rookie, let me show you how it's done." He takes the juvenile into a back room for questioning while you wait outside. You hear blows, grunts, and screams coming from the back room. What will you do now?

Situation 2: You are a patrolman assigned to a new district, and are about to meet your sergeant for the first time. You know that several complaints have been made against you by black citizens, charging abuse and unnecessary use of force. You suspect that your abrupt transfer to this new district is the result. You knock on the sergeant's door. The sergeant is to give you your new assignment for duty.

Situation 3: You are a rookie patrolman on duty. You have discovered a break-in of a fur warehouse. You have reported it to your sergeant, and a team has just completed the investigation. You are sitting in the cruiser with the sergeant. The sergeant says, "You did a good job, rookie. To reward you, you'll find a little present for your wife in the trunk of your car. When you get off duty tonight, look in your trunk. I hope your wife needs a fur coat." What do you do?

Situation 4: You are a Police Cadet who has been getting good reports from your superiors on the job. You like the work. However, you have gotten failing marks in several of your academic subjects in the Police Science program at UC. The Coordinator of Police Science on campus has called you into his office, and announced that you are being placed on academic probation. You know that you will be terminated from the Cadet program if your grades do not improve. What do you say to the Coordinator?

Situation 5: You are on patrol. You get a radio call to investigate a noisy party at a fancy estate on Wm. Howard Taft Rd. A neighbor has complained of the noise. You arrive, and find a loud rock-and-roll band playing on the lawn. You recognize a prominent local judge and several city officials drinking on the terrace. Mr. Martinelli, the host, appears inebriated, and tells you this is a private party, and police were not invited. What do you do?

Situation 6: You are on patrol. You get a radio call to investigate a noisy party in an apartment on Calhoun Street. A neighbor has complained of a stereo being turned up to full volume. You knock on the door, and a bearded student answers. The apartment is crowded with young people, and you think you smell marijuana. The man tells you that this is a private party, and the fuzz can keep out. What do you do?

Situation 7: You are on patrol with a patrolman-coach. You stop into a small restaurant on Vine Street for a hamburger and cup of coffee. Your coach instructs you just to show your ID to the cashier, that police always eat free here, and the owner appreciates the protection. What do you say?

Situation 8: You are on patrol on Vine Street at night. You see an intoxicated man stumble out of Pappy's Bar, and walk down an alley. You recognize him as a man you have arrested previously for public drunkenness. You park your cruiser at the curb, and proceed on foot down the alley. You place the drunk under arrest. As you leave the alley, you see that his friends have also left the Swing Bar, and are standing between you and the cruiser.

Situation 9: A black woman has walked into the district station at night, complaining that her husband beat her up. She appears intoxicated. She refuses to prosecute her husband, fearing that he will "kill her," but insists that she wants police help to "stop him from drinking." Finally the sergeant orders you to go up to this family's tenement apartment "to see if you can straighten this man out." You arrive, and find that the door is locked, and the husband is unwilling to open it.

Situation 10: You are on traffic patrol on Columbia Parkway late at night. A Cadillac speeds by, and you take up pursuit. He is traveling at an excessive rate of speed, and crosses the center line twice. You finally persuade him to pull over, and as you approach the driver, you recognize him as one of the County Commissioners. He is very belligerent and indignant about being stopped.

Situation 11: A store-owner at 8th and State has been complaining to your sergeant about juveniles hanging around on the sidewalk in front of his store. These young men whistle at girls, block the entrance to his store, and he suspects thefts of merchandise by these youths. As you pass by, you see that this gang is again leaning against his store front. You park your cruiser, and approach on foot. A tall, bearded young man detaches himself from the group, and walks forward to meet you.

Situation 12: You have received a trouble call from the Red Barn restaurant on Vine Street. The manager reports that a black youth insulted a waitress, called her "an Appalachian whore." The manager points out the youth who is sitting in his car nearby. The manager says he doesn't want any trouble, but insists that this youth must get off the restaurant's property.

Situation 13: You receive a radio call to investigate a petty larceny at Ontario's. The manager tells you that a young man has been caught walking out of his store without paying for a set of fancy hub-caps, and takes you into his office. You recognize the suspect as Jim Norton, whom you played football with on your high school team last year. You request an opportunity to talk with the suspect privately.

POLICE RECRUIT ROLE-PLAYING SITUATIONS
(LSYTYWOOTO)

INSTRUCTIONS: Divide group into pairs. Each pair is to alternate playing roles prescribed in situations. Entire group is to vote after each "round" on overall effectiveness with which each participant played role, voting for member of pair who was most effective in role. Vote to be recorded, roles switched, vote again taken, and results announced to group. Post on easel. Precaution: No physical combat is permitted in this exercise.

Situation 1: You are a patrolman on duty on foot at night in the downtown district. You hear a burglar alarm ringing. As you round the corner to investigate, you see a suspect standing in front of a broken jewelry store window with his hands full of watches. You approach the suspect. The suspect makes no attempt to flee.

Situation 2: You are a patrolman assigned to a new district. You are meeting your sergeant for the first time. Both the sergeant and patrolman know that bad reports on your conduct have been placed in your personnel record, and you suspect that your transfer to the new district has been the result. You walk into the sergeant's office.

Situation 3: You have received a trouble call from the White Castle restaurant at Reading Rd. and Wm. H. Taft Rd. A waitress states that a Negro juvenile made insulting remarks to her when she served him. She points him out sitting in a car outside. You approach the juvenile's car. He is sitting eating a hamburger.

Situation 4: You are an undercover detective investigating drug abuse along Calhoun Street. A young hippie with long hair, whom you casually know, approaches you. He asks whether you would like to go with him to his apartment.

Situation 5: A store-owner in Avondale has complained that groups of young men "hang around" in front of his store and accost customers occasionally. You notice 6 young men on the sidewalk in front of his store. As you get out of your cruiser, one of the young men, with a beard, detaches himself from the group, and walks toward you.

Situation 6: You are on traffic patrol at night. You see a car weaving in and out of lanes on Columbia Parkway. You flag the driver over to the side of the road. He complies. You get out of your cruiser and approach the driver seated in his car.

Situation 7: A Negro woman comes into the District station. She has obviously been drinking, and has bruises on her face. She states that her husband beat her up, but refuses to sign a complaint. She is afraid to go home, and asks the police to help her. The sergeant asks you to investigate, and you go to the address the woman has indicated, and knock on the door.

Situation 8: Around 8 p.m. a neighbor called the police station to report that 3 children, age 3 through 5, are being neglected by their parents. They are presently at home unattended by an adult. You go to the home to investigate and find the children are there alone - they tell you that they are hungry and there is no food in the house. The house is dirty but shows some signs of being cared for. As you are talking, the father/mother comes in the door and demands angrily what you are doing in his/her home. There is a distinct odor of alcohol on the person.

Situation 9: Recently there has been a sharp increase in shoplifting in the Swifton Shopping Center. Police have had difficulty prosecuting suspects even when they have been caught red-handed. Local community groups and individual citizens are charging that police are indiscriminately picking up black youths for questioning and prosecution. The store detective from Mabley & Carew calls the police station to report the theft of a \$350.00 watch. You interview the store clerk who states that she noticed the watch missing after having shown it to 3 persons. One looked like a middle-class businessman, another a suburban housewife, and the third a black youth of about 16 wearing jeans and a large Afro. As you leave the store you notice 3 black youths matching the description. Two, after receiving a package from the third, disappear into the crowd before you reach the spot. You approach the youth who seems very hostile.

Situation 10: You answer a call for assistance from a playground in the Over-the-Rhine area. The recreational supervisor, an attractive young black woman, reports that a young man has been making insulting remarks to her and also been making agitating remarks "e.g. honkie" etc. to some of the white youths on the playground. She points the young man out. At this time he is busily playing basketball with about 10 other black youths.

Situation 11: A local industrialist is giving a party for a visiting nationally known politician. It is a garden party with about 200 guests, a live band and a singing group. Several neighbors complain about the noise. You are sent to investigate. The mayor, several local judges, university presidents and other well-known persons are present. The host is a little high and very angry at your intrusion and demands that you leave immediately.

THE LEGAL ASPECTS
OF
PSYCHOLOGICAL TESTING

1. Background.

a. On March 24, 1972, Title VII of the Civil Rights Act of 1964 was amended by the Equal Employment Opportunity Act of 1972 thereby applying across the board to Federal, state and local public safety employers. The purpose of Title VII is to achieve equality of employment opportunities and remove barriers that have operated in the past in favor of an identifiable group of white employees over other employees.¹ The act now is applicable to any public safety employer who employs 15 or more employees for each working day in each of 20 or more calendar weeks. The Title also covers labor organizations. Under the statute, labor organizations are defined as any organization, agency employee representation committee, association, etc., in which employees participate and whose purpose is to deal with employers concerning grievances, wages, hours or any other conditions of employment covered by the Act.

b. The commerce clause of the Constitution establishes jurisdiction over state and local employers and unions.² Title VII applies where

1. Griggs v Duke Power Co., 401 US 424, 429 (1971).

2. United States Constitution Article 1, Section 8, Clause 3.

employer or labor is involved in an industry affecting commerce, and Section 701(g) of the Act defines industry affecting commerce to include any governmental industry. Title VII does not apply to public officials, their personal staff and appointees on the policy making level or immediate advisors with respect to the constitutional or legal power of the elected officials unless his job is subject to state or local Civil Service laws.

c. The following unlawful employment practices are prohibited by Title VII.

(1) Discrimination in hiring, discharging, compensation, or in terms, conditions, or privileges of employment based on an individual's race, color, religion, sex, or national origin.

(2) Limiting, segregating, or classifying employees in such a fashion as to adversely affect the employee's employment opportunities because of race, color, religion, sex, or national origin.

(3) Discrimination by a labor organization against an individual based on race, color, sex, religion, or national origin.

(4) Discrimination in training or retraining programs, including on-the-job training programs, based on an individual's race, color, religion, sex, or national origin.

(5) Discrimination against an employee or labor organization member for opposing any practice made unlawful under Title VII.

(6) Including in any notice or employment advertisement any preference, limitation, specification, or discrimination based on race, color, religion, sex, or national origin.

2. Discussion.

a. The new law creates certain exceptions to practices which would otherwise be unlawful. The following exception is relevant to the area of psychological testing:

Section 703(h) states that it will not be an unlawful practice to "give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex, or national origin." The courts and the EEOC will be looking for the discriminatory consequences of testing. If the test has a discriminatory effect, the employer will have a difficult burden in trying to show the reasonableness and necessity for the test.

b. Section 703(h) of Title VII permits employers to use professionally developed tests provided that such tests, their administration or

action upon the results are not designed, intended, or used to discriminate because of race, color, religion, sex, or national origin.³ The Supreme Court in Griggs interpreted Section 703(h) under Title VII to hold the following:⁴

(1) Title VII does not command that any person be hired simply because he was formally the subject of discrimination, or because he is a member of a minority group.

(2) The Act was intended to remove artificial employment barriers that discriminate on the basis of racial or other impermissible classification.

(3) The Act prohibits not only overt discrimination but also practices that are fair in form but discriminatory in effect.

(4) If an examination has as its consequence the exclusion of a group protected under Title VII, the test will be unlawful unless the employer can prove it to be job-related.

(5) The EEOC guidelines on test validation are to be given great deference. Holdings 4 and 5 above have the greatest impact on the use of

3. Griggs v Duke Power Co., 401 US 424, 429 (1971).

4. IBID.

psychological testing in law enforcement selection. If an examination disproportionately excludes a class of persons (i.e. blacks, women), the employer must satisfy the EEOC employment testing guidelines on examination validation. Tests are defined under the guidelines as any performance measure used as a basis for an employment decision. This definition includes such items as personal history/background examination, specific education or work history requirements, scored interviews, biographical information blanks, scored application and interviewers rating scales. Any tests must be demonstrated by the employer to be valid predictors of employee job performance (validation). This validation can be achieved in one of two ways:

(a) Empirical validation showing strong correlation between tests scores and subsequent job performance.

(b) Content construct validation involving an analysis of the tests to be performed in the job in question on an examination built around the job analysis.

c. The legal sufficiency of the paper and pencil psychological tests used in the selection of law enforcement personnel leaves much to be desired. All tests reviewed to date (with one exception discussed in

d below), including the use of the Army ST score,⁵ have the following deficiencies:

- (1) They measure and predict only scholastic performance.
- (2) They are not built on a validated analysis of law enforcement job performance.
- (3) They do not deal with the conclusive evidence that there are racial differences in test responses.

By now it is becoming common knowledge that written general ability examinations currently prevalent as patrolmen employment tools measure only the applicants' capabilities in the abstract and not prediction of job performance. However, such examinations have been found to discriminate against minority group applicants seeking careers in law enforcement.

d. In April of 1972 the Center of Criminal Justice Agency organization and minority employment opportunities of Marquette University Law School addressed the problem of police manpower selection. In the search for valid police entry tests that are culturally fair, the Marquette

5. Effectiveness of Selection and Classification Testing, US Army Research Behavioral and Social Sciences, Sep 73.

Center found the field to be barren.⁶ The Industrial Relations Center of the University of Chicago sponsored by the United States Department of Justice produced strong evidence for the validity of the tests they employed in the selection of law enforcement personnel. The results of the 1969-71 follow-up study verified the predictive validity of the test battery over time for both the supervisory measures of patrolman success (i.e. the paired comparison ratings and CPD rating) and for a number of the more objective measures of performance. Predicted performance levels, based on test scores obtained in 1966, proved to have statistically significant and meaningful correlation with subsequent job performance in 1967-72. The second result of the validation analysis indicated that it was necessary that separate validations were needed for different racial groups. Separate validations produce selection procedures which apply the same standards of performance to all applicants but take into account the fact that different patterns of test scores can predict the same level of performance. The result is a possible increase in the number of minority applicants who would be accepted and perform under this system when they would fail other systems not designed to assess their ability to perform on the job.

6. Personal Communication, Charles Mentkowski, Associate Dean, Marquette University Law School.

e. Since the Griggs, et al. versus Duke Power Supreme Court case (1971), which affirmed the EEOC guidelines on employee selection and promotion, organizations must be prepared to offer proof that their selection and appraisals devices, including assessment centers, are valid and job-related.

f. Properly developed assessment centers can be shown to be job-related through their content. If the variables measured by the center were arrived at through an accurate and complete job analysis; and if the exercises and tests used accurately measure these variables, the procedure is valid and job-related. Center exercises often are "job samples" of the kind of work managers actually perform, just as a typing test is a sample of a typist's job, and thus possess rational validity; they make good common sense. An In-Basket exercise obviously qualifies as a managerial job sample. An exercise measuring group effectiveness can also be considered a job sample if given to potential executives who will spend a great deal of their time in meetings. Perhaps because of the "rational validity" of the assessment center approach, no known charges of discrimination resulting from application of an assessment center have been filed anywhere in the United States.

g. There is strong statistical evidence from organizations experienced with the method that the assessment center procedure is, in general, extremely valid when applied to minority group members. Moses (in press)

examined both one-day and three-day assessment programs at AT&T and reported that no differential validity results were found between any of the subgroups based upon race and/or sex and assessment performance. The only other published study of differential validity was done by Meyer (1972) and dealt with assessment centers at General Electric. Of 120 whites assessed, 25% received job offers while of 30 blacks assessed, 50% received offers. Thus, an indirect bias in favor of blacks appeared to be operating. While validity in one organization does not necessarily mean that the procedure is valid in another, the existence of these studies would be an important consideration in any court case involving assessment centers.

h. The implications are not that all assessment centers should be considered valid. Any organization adopting the center approach should establish empirical procedures to evaluate and analyze data. Relationships between assessment center performance and job effectiveness criteria must be thoroughly investigated. In addition, some procedure should be established to see that all potential assessees are given an equal opportunity to attend the center.

3. Conclusions:

a. None of the pencil and paper psychological tests employed as a screening technique in the selection of law enforcement officers have been blessed by the courts.

b. Although the assessment center approach appears to be more easily validated for court approval, only the Mills study is completely tailored toward law enforcement personnel selection.

c. The use of a minimum ST score by the US Army in the selection of military police has been untested in its validity to select police personnel. It is not job related and appears to be discriminatory toward minority personnel.

d. Although not validated to court acceptance, the test battery structured by the Industrial Relations Center of the University of Chicago shows both job related and ethnic equality potential.